

The author hereby certifies that the use of any copyrighted material in the dissertation entitled:

“Validity of Retrospective Reports of Eating Behavior from the Eating Disorder Examination”

beyond brief excerpts is with the permission of the copyright owner, and will save and hold harmless the Uniformed Services University of the Health Sciences from any damage which may arise from such copyright violations.

ABSTRACT

Title of Thesis: Validity of Retrospective Reports of Eating Behavior from the Eating Disorder Examination

Capt. Jay M. Stone, Master of Science, 1999

**Thesis directed by: Tracy Sbrocco, Ph.D.
Assistant Professor
Department of Medical and Clinical Psychology**

The Eating Disorder Examination (EDE; Cooper & Fairburn, 1987) is the most widely used instrument for the diagnosis of eating disorders and relies on retrospective self-report. However, there is growing evidence that retrospective self-reports are prone to errors arising from autobiographical memory. Stone and Shiffman (1994) adopted a method for collecting moment-by-moment data to address these concerns. The present study examined the accuracy of these estimates by comparing retrospective reports from questions on the EDE with data recorded in handheld computerized eating diaries by obese and normal-weight women.

The results suggest some lack of correspondence between the diary data and the EDE for a frequency count of most meal types and for overeating days and episodes, as well as for most cognitive-affective states. Many responses on the EDE appeared anchored at either end, reflecting endorsements of daily or never. However, moment-by-moment recording in the eating diary reflected a range of responses.

**VALIDITY OF RETROSPECTIVE REPORTS OF EATING BEHAVIOR FROM THE
EATING DISORDER EXAMINATION**

by

Jay M. Stone

**Thesis submitted to the Faculty of the
Department of Medical and Clinical Psychology Graduate Program of the
Uniformed Services University of the Health Sciences
In partial fulfillment of the requirements for the degree of
Master of Science 1999**

Table of Contents

List of Tables

List of Figures

Background

The Eating Disorder Examination

Cognitive Science Perspective of Dietary Recall

Ecological Momentary Assessment

Problems with Self-Monitoring Food Intake

Purpose of Present Study

Method

Participants

Measures

Eating Disorder Examination

Eating Records

Procedure

Data Reduction and Analyses

Results

Discussion

References

Tables

Figures

List of Tables

Table 1. DSM-IV Diagnostic Criteria for Bulimia Nervosa

Table 2. DSM-IV Diagnostic Criteria for Anorexia Nervosa

Table 3. DSM-IV Eating Disorder Not Otherwise Specified

Table 4. DSM-IV Research Criteria for Binge Eating Disorder

Table 5. The Four Subscales of EDE 12.0D

Table 6. The EDE Rating Scheme

Table 7. Obese and Normal Weight Participant Characteristics

Table 8. Comparisons of reported meal frequencies for the EDE and handheld computer eating diary

Table 9. Comparisons of the EDE and handheld computer eating diary assessment of restraint over eating

Table 10. Comparisons of the EDE and handheld computer eating diary assessment of avoidance of eating

Table 11. Comparisons of the EDE and handheld computer eating diary assessment of days of overeating

Table 12. Comparisons of the EDE and handheld computer eating diary assessment of episodes of overeating

Table 13. Assessment of episodes of overeating from the handheld computer eating diary

Table 14. Comparisons of the EDE and handheld computer eating diary assessment of reaction to prescribed weighing

List of Figures

Figure 1. Proposed Model for the Storage and Retrieval of Information about Frequent Behavior (Menon, 1993)

Figure 2. The EDE Scheme for classifying episodes of overeating

Figure 3. Obese' Meal Frequency Distributions

Figure 4. Normals' Meal Frequency Distributions

Figure 5. Restraint Over Eating Frequency Distributions

Figure 6. Avoidance of Eating Frequency Distributions

Figure 7. Reaction to Prescribed Weighing Frequency Distributions

The assessment of eating disorders and problem eating behavior relies on the retrospective report of behavior, including frequency counts of overeating, purging, fasting, excessive exercise, and restrained eating. The Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV; APA, 1994) describes three categories of eating disorders: Bulimia Nervosa, Anorexia Nervosa, and Eating Disorder Not Otherwise Specified. In order to make these diagnoses, clinicians must rely heavily on the validity of patients' retrospective self-report. Though it is well recognized that self-report, particularly retrospective self-report, is suspect, there has been little examination of this method of assessment in clinical diagnosis. This is particularly surprising given the DSM IV's sole reliance on this method of assessment. Several of the key features among the Eating Disorders rely on retrospective report and frequency counts of behaviors. The major diagnostic interview used to assess Eating Disorders relies on retrospective self-report. The full DSM-IV criteria for Bulimia Nervosa, Anorexia Nervosa, Binge Eating Disorder and Eating Disorder NOS are presented in Table 1 (see Tables 1 - 4).

Recurrent episodes of binge eating, a major criterion for Bulimia Nervosa, is assessed by asking the patient to estimate the number of binge eating episodes they have experienced in the past three months, as well as the amount of food consumed during these episodes. A similar assessment of binge eating is required for Binge Eating Disorder over an even longer (6 month) time period. Similarly, assessment of "inappropriate compensatory behaviors" that characterize Bulimia Nervosa requires retrospective frequency counts of self-induced vomiting, misuse of laxatives or diuretics, fasting, and excessive exercise.

To date, only three semi-structured assessment instruments have been developed and validated to assess the DSM-IV criteria needed to make clinical diagnoses for eating disorders: the Eating Disorder Examination (EDE; Cooper & Fairburn, 1987; Fairburn & Cooper, 1993), the Clinical Eating Disorder Rating Instrument (CEDRI; Palmer, Christie, Cordle, Davies, & Kendrick, 1987) and the Interview for Diagnosis of Eating Disorders (IDED; Williamson, 1990).

Of the three interviews, the EDE has received the greatest attention from researchers (Williamson, Anderson, & Gleaves, 1996).

A number of self-report instruments have also been designed to assess the different features of disorder eating pathology. While diagnosis cannot be inferred solely on the basis of self-report, cut-off scores on these measures have been established to differentiate individuals with eating disorder pathology. These instruments include the Eating Attitudes Test (EAT; Garner & Garfinkel, 1979), the Eating Disorder Inventory, Version 2 (EDI-2, Garner, 1991), the Binge Eating Scale (Gormally et al., 1982), the Bulimia Test (Smith & Thelan, 1984), the Bulimic Investigatory Test (Henderson & Freeman, 1987), the Revised Restraint Scale (Herman & Polivy, 1980), the Three Factor Eating Questionnaire (Stunkard & Messick, 1985), the Dutch Eating Behavior Questionnaire (van Strien et al., 1986) and the Body Shape Questionnaire (Cooper et al., 1987). Although they possess acceptable psychometric characteristics and are useful in many situations, there are limitations of this type of self-report questionnaire.

Many of the questionnaires are inadequate. Some of the important problems are: (a) the rating scales for certain critical behaviors, e.g., binge eating and vomiting, consist of vague verbal descriptors of severity such as “never”, “often”, “always”, etc.; (b) there is no time frame for reports of symptoms on most questionnaires; (c) there are no objective definitions for behavioral symptoms of binge eating or dieting; and (d) there is no means to discriminate complaints of truly pathologic body image and eating attitudes and behaviors from those which may be subjectively distressing, but which are widely held in our weight-conscious society. Self-report on questionnaires can be ambiguous and inaccurate (Rosen & Srebnik, 1990). As it currently stands, eating disorder diagnoses cannot be made from questionnaire data.

The Eating Disorder Examination

Many have argued that the clinical interview is the best methodology to obtain the fine-grained detail necessary for the assessment of eating disorders (Cooper and Fairburn, 1987;

Palmer, Christie, Cordle, Davies, & Kendrick, 1987; Williamson, 1990). The EDE, developed by Cooper and Fairburn (1987) is the most widely used diagnostic instrument for the diagnosis of eating disorders. The interview normally takes 30 – 45 minutes to complete. The EDE was developed to provide standard definitions of critical symptoms of eating disorders. The most recent version was developed to assess DSM-IV criteria for eating disorders. Among the unique features of the EDE, the interviewer rates the participant's abnormality of eating rather than relying upon the participant's own terminology. The instrument has been progressively refined over the past 10 years to maximize its reliability and validity and is now in its 12th edition. In its original form, the EDE was designed to assess present state and as such focused exclusively on the previous 4 weeks. In this form the instrument generated basic descriptive information on the degree of behavioral disturbance (e.g., frequency of various forms of overeating, self-induced vomiting, etc.) as well as a profile of individuals in terms of their scores on five subscales designed to assess key aspects of eating disorder psychopathology (Bulimia, Restraint, Eating Concern, Shape Concern, and Weight Concern). More recently a diagnostic version of the EDE has been developed that also generates operationally defined eating disorder diagnoses from the DSM-IV (EDE 12.0D; Fairburn & Cooper, 1993). As a result, certain features of diagnostic importance are assessed over a 3-month period.

In addition to allowing for DSM-IV eating disorder diagnoses, the EDE provides three levels of descriptive data concerning current eating disorder psychopathology: scores on individual items, subscale scores, and a global score. The EDE provides either frequency or severity ratings (0 – 6) for key behavioral and attitudinal aspects of eating disorders (see Table 6). In the case of certain frequency scores (e.g., those for bulimic episodes and self-induced vomiting), it may be difficult to obtain an accurate estimate of the rate of the behavior especially if it is occurring very frequently. In such cases it is recommended that frequencies are reported for the number of days on which the behavior occurred. The EDE assesses two key behavioral aspects of eating disorders and provides frequency ratings for their occurrence. These are

overeating and the use of extreme methods of weight control. Three forms of overeating (objective and subjective bulimic episodes, and episodes of objective overeating) are measured both in terms of their absolute frequency and the number of days on which they occurred. Four extreme methods of weight control (self-induced vomiting, laxative misuse, diuretic misuse, and intense exercising) are assessed. Some of the individual items are used to make DSM-IV eating diagnoses. These include Bulimic Episodes and Overeating, Dietary Restriction Outside Bulimic Episodes, Self-Induced Vomiting, Laxative Misuse, Diuretic Misuse, Intense Exercising to Control Shape or Weight, Abstinence from Extreme Weight Control Behavior, Importance of Shape, Importance of Weight, Fear of Weight Gain, Feelings of Fatness, Maintained Low Weight, and Menstrual History.

Subscale scores provide a profile of individuals in terms of four major areas of eating disorder psychopathology. A set of five subscales was originally derived from the EDE by grouping items together to represent the major areas of specific psychopathology (Cooper, Cooper, & Fairburn, 1989). This rational assignment of items to subscales was checked empirically by examining the internal consistency of the subscales and, as a result, certain minor adjustments were made. Four of these original subscales remain unchanged in EDE 12.0D. The Bulimia subscale has been omitted because it does not add further descriptive information beyond that which can be derived from the frequencies of the various forms of overeating. Listed in Table 5 are the current subscales and the items comprising them. To obtain a particular subscale score, the ratings for the appropriate items are added together and the sum divided by the total number of items forming the subscale.

The global score provides a measure of the overall severity of the eating disorder psychopathology (e.g., Fairburn, Peveler, Jones, Hope, & Doll, 1994). To obtain a total score on the EDE the subscale scores are summed and the resulting total divided by the number of subscales (i.e., four). It is recommended that the global score be reported in conjunction with detailed EDE data for both individual subscales and key behavior.

The EDE has been used in descriptive studies (e.g., Beumont, Kopec-Schrader, Talbot, & Touyz, 1994; Marcus, Smith, Santelli, & Kaye, 1992; Taylor, Peveler, Hibbert, & Fairburn, 1994; Wilson & Smith, 1989) and research on treatment (e.g., Fairburn, Jones, Peveler, Hope, & O'Connor, 1994; Garner et al., 1993; Wilson, Eldredge, Smith, & Niles, 1991). Adaptations have been devised for those who are pregnant and those with diabetes mellitus (Fairburn, Peveler, Davies, Mann, & Mayou, 1991; Fairburn, Stein, & Jones, 1992; Peveler, Fairburn, Boller, & Dunger, 1992; Striegel-Moore, Nicholas, & Tamborlane, 1992).

The EDE reportedly has favorable psychometric properties. High interrater reliability has been demonstrated in different settings (Cooper & Fairburn, 1987; Rosen, Vara, Wendt, & Leitenberg, 1990; Wilson & Smith, 1989). The internal consistency of the five subscales is satisfactory (Cooper, Cooper, & Fairburn, 1989). Its discriminant validity, as a measure of the specific psychopathology of eating disorders in general, and binge eating in particular, was established in a study of 100 patients with anorexia nervosa or bulimia nervosa versus normal controls (Cooper et al., 1989). All individual items showed significant differences between the two groups. The selection criteria for the control group in this study quite possibly resulted in the inclusion of participants who were unconcerned about body weight or shape. However, Wilson and Smith (1989) showed that the EDE distinguished between patients with bulimia nervosa and nonbulimic participants who were preoccupied with dieting and weight. In this study the Eating Disorder Inventory (EDI) subscales did not discriminate between the two groups, indicating the superior discriminant validity of the semistructured interview to this standardized self-report questionnaire. The advantage of the EDE over the EDI appears to have been the freedom of the interviewer to pursue information in greater detail with the EDE. However, this certainly does not discount the validity of self-report surveys, such as the EDI, as they have repeatedly proven themselves useful in helping to diagnose a variety of disorders. Finally, the EDE has proved to be a sensitive measure of the effects of psychological treatment of bulimia nervosa (Fairburn et al., 1991; Garner et al., 1993; Wilson, Eldredge, Smith, & Niles, 1991).

Validity studies of the EDE have used daily self-report of eating as a comparative standard. Using a sample of unselected college students, Rosen and colleagues (1990) correlated self-report of all food and liquid intake for the 7 days prior to administration of the EDE with the different subscales. The eating record forms were divided into spaces for breakfast, lunch, dinner, and morning, afternoon, and evening snacks. In addition, spaces were provided for the participants to indicate whether they regarded the eating episode as a binge and whether they had vomited afterwards. The overeating subscale was significantly correlated with frequency of binge-eating episodes. The investigators state that they derived their estimate of binge eating from the students' self-report. Any attempt to derive a count of binge eating from self-monitoring must necessarily rely on participants' interpretation because independent raters cannot infer loss of control, even though they can estimate amount of food consumed. However, unless the students were given a definition of binge eating (which is not made clear), it can be assumed that they reported both subjective and objective bulimic episodes as "binges." The correlation between the EDE and students' self-recordings, therefore, may actually underestimate the degree of concordance between the two forms of assessment.

Rosen and colleagues (1990) concluded the EDE has moderate concurrent validity with measures of dietary restraint and overeating taken from eating records, including average calorie intake, avoidance of regular meals, avoidance of snack food or forbidden foods, frequency of binge eating, and size of binge episodes. In addition, EDE ratings of vomiting showed good agreement with self-monitored vomiting on diaries. Thus, these data suggest that the EDE is a valid, albeit far from exact, measure of eating behavior.

However, the findings of Rosen and colleagues (1990) are limited by the fact that only subscale scores of the EDE were compared with diary data, not scores on individual items. It is therefore still unknown how well many EDE individual items (e.g., frequency of meals) correlate with daily diary entries. The findings are also limited by the fact that only bulimia nervosa participants and "dietary-restrained" control participants were used in the study.

In an extension of this research, Loeb, Walsh, and Pike (1992) compared EDE assessment of binge eating (objective bulimic episodes) with self-recording in bulimia nervosa patients at pre- and posttreatment. At pretreatment, the correlation between the two measures for the identical preceding seven days was an impressive .96. At posttreatment, Loeb and colleagues (1992) found that the two measures for the same seven and 28-day periods were .97 and .98 respectively. Such high correlations might well reflect patients' memories of self-recording binge frequencies during the preceding 7 and 28 days. As a result, the data do not directly address the question of the accuracy of recall on the EDE for the preceding time period in the absence of formal self-monitoring of binge eating. For example, we do not know for sure what memory processes are used by respondents to answer questions from the EDE such as "Over the past 4 weeks which of these meals or snacks have you eaten on a regular basis?" and "I would like you to describe any times when you have felt that you have eaten too much in one go. What were others eating at the time? Did you have a sense of loss of control at the time? Could you have stopped eating once you had started?" These kinds of specific questions appear to be difficult for anyone to recall accurately. Knowing what memory processes are used by respondents to answer these kinds of questions is important for two reasons: 1) it gives the evaluator an idea of how much credibility to give the responses, and 2) it provides a framework from which techniques can be used to enhance recall.

While we regularly rely on individuals to provide quite complex accounts of past behavior, we rarely question their ability to do so or the method by which they "gather" this information. For example, how does an individual remember how many times they binged over the past month? Do they sit down and recall each day, one at a time? Do they think to themselves, "Well, I have binged three times during the past week and that is pretty typical for me so I'll "guestimate" that I've binged 12 times in the past month." Do all individuals use the same method? Does the method depend on how regularly the behavior occurs? If we do stop to think about what we ask individuals to do, it seems clear that in many cases we are getting

estimates of behavior. Therefore, the accuracy and methods by which individuals recall the information we rely on to make assessments and develop treatment plans is obviously important. Fortunately, the fields of survey research and cognitive science provide us with some useful information to help us address these problems.

Cognitive Science Perspective of Dietary Recall

Concerns about the accuracy of food recall in epidemiological research has led to the application of cognitive science paradigms to examine the accuracy of dietary recall. During the past 15 years, investigators in the fields of survey research and cognitive psychology have begun interdisciplinary work to improve survey methodology (Jobe & Mingay, 1991). Much of this research has been devoted to the accuracy of dietary recall. The majority of studies concerning this issue have detected bias in self-reported food intake (Schoeller, 1990), with reported food intake typically underestimating actual food intake (Johnson, Goran, & Poehlman, 1994; Schoeller, 1990). Combined findings from a number of studies indicate that there is poor correspondence between self-reported dietary intake and total energy expenditure (Bandini, Schoeller, Cyr, & Dietz, 1990; Johnson et al., 1994; Schoeller, 1990).

Smith, Jobe, and Mingay (1991) conducted two studies to obtain descriptive information about dietary recall in order to hypothesize a theoretical characterization of this task. In the first study, participants recorded their diets for 2 or 4 weeks, and returned 0, 2, 4, or 6 weeks after the end of the recording period for a memory test which required that they attempt to report all items that they had eaten or drunk during the recording period. The reported items were scored against the recorded ones. In the second study, participants recorded their dietary intake for two separate periods, and reported for both periods at the end of the second.

The results indicate that specific memories do contribute to long-term dietary reports. Memory performance deteriorated as a function of retention interval, and this performance decline was manifested both as a decreasing target report rate and as an increasing intrusion rate.

The investigators attribute this to the decay of specific memories from the reference period (cf. e.g. Ebbinghaus, 1885/1964). However, other aspects of the data suggest that participants' reports are not based exclusively on specific memories, but rather rely substantially on generic memory.

First, that intrusion rates were substantial even at the shortest retention interval indicates that participants do not base responses exclusively on their specific dietary experiences during the reference period. The reported items that the researchers scored as intrusions were presumably items that these participants ate at some time. Intuition suggests that these items are reported on the basis of generic knowledge rather than by misdating specific remembered episodes, although their studies did not address this.

Second, over increasingly long retention intervals the intrusion rates of participants who reported about the shorter reference period increased, whereas those of participants who reported about the longer period were approximately constant. This pattern of intrusions would result naturally if all participants simply described their typical diets. Because a 4-week period is more likely than a 2-week period to contain any item that an individual routinely eats, participants who reported for a 4-week period were considerably more free to report items without concern for when they were eaten.

Third, because food groups are a plausible basis of organization of general knowledge about food and personal diet, the effect of reporting guidelines on performance was consistent with what one would expect if participants relied on such knowledge when they reported. Reporting by food groups elevated the number of items reported by participants, but did not improve the match rate. Deese (1961) argued that all free recall involves production of a response list based on associations to a few retained elements; construction of such a list is, perhaps, especially feasible when all of the to-be-reported items come from a single category.

Finally, within-subjects data, collected in the second study, showed that the items reported for a remote period matched what was eaten during that period no better than they

matched what was eaten during a recent period, and that both of these pairings revealed a modest match rate and a fairly high intrusion rate. It appears that reporting a generic set of items would match, at least roughly, what was eaten during any period. The contribution of specific memories to dietary reports for recent periods must be evaluated as an increment in performance over the baseline level that would be obtained by reporting a generic set of items. This baseline score is substantially higher for reports of dietary intake than that which might be obtained in ordinary free recall by reporting a “generic” set of words (cf. Erdelyi, Finks, & Feigin-Pfau, 1989).

The Smith, Jobe, and Mingay (1991) findings are compatible with others in the literature on memory. For example, Graesser (1981) found, in experiments with carefully controlled stimulus materials, that the importance of memory schemas in guiding retrieval increases as retention interval lengthens. At short retention intervals (such as daily recording of eating behavior), memory for schema-atypical items was better than that for schema-typical items but, over time (such as recalling eating behavior over the past 30 days on the EDE), memory for atypical items decayed more rapidly than memory for typical ones. Graesser argued that, by some point in time, virtually nothing will be remembered about the to-be-reported events, and that, after that point, participants will respond entirely on the basis of the relevant schema. Reder (1987a,b) distinguished between direct retrieval and plausible inference as strategies for question-answering, and showed that as the amount of time since the learning of information relevant to a question increased, plausible inference supplanted direct retrieval as the preferred strategy of question-answering. The data on intrusions in free recall of lists of categorically related words make the same point (e.g. Cofer, 1967; Rabinowitz et al., 1977). The Smith and colleagues (1991) data are consistent with this general pattern of results.

Smith, Jobe, and Mingay (1991) point out that although these studies focused on dietary recall, they expect that their pattern of findings and their account of them would be appropriate to describe memory-based reports for other variegated classes of events that occur frequently and repetitively. In particular, they expect that this account characterizes survey respondents who,

like experimental subjects, probably answer questions about such classes of events without remembering the specifics of their experiences.

Friedenreich (1994) proposes a 4 stage model describing the cognitive processes undertaken by respondents when answering a question. These stages are question comprehension, information retrieval, estimation and judgment, and response formulation. During the first stage, the respondents interpret the meaning of the question. In the second stage, they search long-term memory for the relevant information. The third stage occurs when respondents evaluate the information retrieved from memory and decide whether it is relevant and adequate. At this stage, if the respondents decide the information is adequate, a response is formulated. Alternatively, if the information is deemed inadequate, they may initiate another search of their memory. In the final stage, the respondents decide what answer to provide by weighing several factors. Although described here as a sequential process, the entire sequence has been hypothesized to be quite flexible, with numerous control processes (decision and judgment) occurring before and after retrieval of information from memory (Willis, Royston, & Bercini, 1991).

At each of these four proposed stages of the cognitive process, the possibility exists that personal and methodologic factors will introduce reporting errors and bias (Friedenreich, 1994). During question comprehension, factors such as age, sex, education, intelligence, ethnicity, disease status, and personal experiences could influence the ability of the respondent to answer the question (Means, Swan, Jobe, & Esposito, 1991). When retrieving the information from memory and evaluating whether it is correct, the determinants of recall ability are likely to be the time interval since the event, the type of information being recalled (episodic or generic information), the amount of detail in the question, the salience of the subject matter to the respondent, the length of the reference period, and the frequency and regularity of the target experience or exposure (Jobe, Tourangeau, & Smith, 1995). Finally, during response formulation, factors that are weighed by the respondent will include the sensitivity of the

question, the social desirability of a particular response, the perceived “correct” response, and the probable accuracy of the answer (Jobe & Mingay, 1989).

Blair and Burton (1987) found that the cognitive processes that respondents use vary depending on the relative frequency of the event. In other words, although it is easy to recall and count every instance for an infrequent behavior, it becomes more difficult to do so for a frequent behavior. Currently, many researchers now maintain that in a survey situation in which respondents are asked a question relating to the frequency of a fairly frequent, nonsalient behavior, they do not do a straight-forward recall and count of every occurrence of the target behavior. Instead, they provide an estimate based on various inference strategies (Blair & Burton, 1987; L. Ross, 1984; Schwarz, 1990; Strube, 1987).

Menon (1993) posits that while the irregularity of the behavior determines the inaccessibility of a ready rate of occurrence as a basis on which to make a behavioral frequency estimate, the similarity of the behavior determines the location of the more accessible information as the semantic or episodic-semantic stores. Figure 1 illustrates the store that Menon’s (1993) proposed autobiographical memory model predicts would be tapped in order to arrive at a frequency estimate of the target behavior.

The figure illustrates that if the behavior is a regular-similar (R-S) one, the procedure that the respondent will use to report a behavioral frequency is straightforward rate-based estimation for the time period of interest to the researcher. In addition, a few minor adjustments may be made on the basis of a recent occurrence or nonoccurrence of the event. The primary accessible store, however, is the semantic store.

If the behavior is a regular-dissimilar (R-DIS) one, the accessible store is likely to be episodic in nature (because of the dissimilarity of the behavior) but with a rate of occurrence (because of the regularity of the behavior) that the respondent can easily resort to in arriving at a frequency. However, because the dissimilarity of the behavior may cause some episodes to be more salient than others, there may be a tendency for respondents to use these available episodes

to make adjustments to the frequency estimate. In addition, the dissimilarity of the behavior may prompt some respondents to use a recall-and-count strategy.

For an irregular-similar (IRR-S) behavior, the more accessible store is still going to be a semantic store because of the homogeneity of the behavior. However, there will be no readily accessible rate of occurrence stored in this semantic store because of the irregularity of the behavior, and the respondent will need to use more cognitive effort in order to estimate the behavioral frequency using some strategy other than a general rate-based one. For example, the respondent may estimate by decomposing the behavior into time and/or situational subdomains and using rates within each subdomain when this is possible (i.e., when there is regularity within subdomains); or the respondent may have to search memory much more to arrive at “available” episodes, which are less accessible for this kind of behavior, and use this as a basis on which to compute a frequency judgment. This process is going to be extremely arduous for the respondent, however, since the accessible store is a semantic one.

Lastly, for irregular-dissimilar (IRR-DIS) behavior, information is likely to be maintained in an episodic format with no general rate of occurrence. The heterogeneity of the behavior is likely to be high, and no general rate of occurrence will be available to which the respondents can resort. They will have to compute behavioral frequency judgments on the spot based on the episodes that are accessible to them, using a counting strategy.

The implications of the Menon (1993) model is that the cognitive effort required by the respondents for behavioral frequency questions depends on the nature of the behavior. Whereas R-S behavior requires the least cognitive effort because of the highly accessible rate of occurrence, R-DIS behavior requires a little more effort, given that some episodes are likely to be still accessible. IRR-DIS behavior requires still more effort, given that there is no ready single rate of occurrence that the respondent has available to compute a frequency judgment. The respondents will have to rely primarily on accessible episodes to compute a frequency judgment directly. Finally, the most complex processes are predicted to be for IRR-S behaviors, which

have very low accessibility of episodes combined with an absence of the general rate of occurrence. The cognitive effort is likely to be the highest in this last case.

Menon's (1993) model would suggest that recalling irregular-similar behaviors, such as eating snacks and episodes of binge eating, requires the greatest cognitive effort because it involves the most complex process of memory retrieval. The results of the Smith, Jobe, and Mingay (1991) studies suggest that respondents rely on generic memory about their own diets when they attempt to report their intake for extended or remote periods (i.e., 2 or 4 weeks). The key questions are, how accurate is accurate enough? and when is specific memory necessary and when can generic memory suffice?

It is difficult to investigate the accuracy of self-report of binge eating and purging because the behaviors are characteristically carried out in secrecy. The use of collateral reports by significant others in the person's natural environment is hence less of an option than it is with other more public forms of psychoactive substance abuse, such as alcohol abuse (Vuchinich, Tucker, & Harllee, 1988). In a rare case in which collateral reports were possible, Giles, Young, and Young (1985) found that they tended to validate the patient's self-report. At best, collateral reports can provide only indirect information on the course of patients' binge eating and purging. In patients with bulimia nervosa, for example, a major goal of treatment is to establish a regular pattern of three nutritionally balanced meals a day. The pattern and content of meals are publicly observable, and in many instances collaterals will have direct and consistent access to this behavior in patients. This sort of collateral report provides indirect but nonetheless useful corroborative information about a patient's progress.

Self-report of binge eating can also be compared with objective assessment of instructed binge eating in hospitalized patients. Mitchell and Laine (1985) found that the average amount of food in binge-eating episodes in the hospital setting was much more than what the patients claimed they normally ate. This does not necessarily indicate that their self-reports were

unreliable. Equally plausible is the interpretation that eating in this artificial setting is unrepresentative of behavior in their natural environment (Rosen & Srebnic, 1990).

There is no biological marker that can be used to validate self-reported food intake patterns, as noted above. More direct measures of energy intake are available and can be used to validate self-reported quantity of food intake. However, these methodologies do not provide information on the size or frequency of meals. The ideal strategy would be to use direct methods of energy intake estimation in combination with self-monitoring and retrospective self-report. Because direct estimations requires more sophisticated laboratory procedures than are available to most individuals, the most feasible strategy to understand the accuracy of retrospective self-report is to compare patients' self-reports of disordered eating with daily recordings of all eating behavior. Loeb and colleagues (1992) and Rosen and colleagues (1990) used this strategy and suggested that the EDE does provide a valid assessment of binge eating in patients with bulimia nervosa. However, the findings are limited by the fact that diary data were not compared with individual items from the EDE (other than bingeing and vomiting in the Loeb et al. study). Rather, comparisons were made with the diary data and subscale scores from the EDE. In addition, these studies did not examine the eating behavior of control participants. There is also data to suggest that there is a marked discrepancy between daily recordings and self-report by obese binge eaters (Rossiter et al. 1992). Whether the difference between these studies is a function of the type of eating disorder, the superiority of the EDE over simple self-report, or other methodological differences remains to be determined.

Ecological Momentary Assessment

Because of the growing evidence that retrospective self-reports are prone to serious errors and biases arising from the characteristics of autobiographical memory, researchers have adopted a variety of methods for collecting moment-by-moment data in real-world settings to address these concerns. One of these methods has been labeled "ecological momentary assessment"

(EMA; Stone & Shiffman, 1994). The hallmark of EMA is the collection of repeated momentary assessments from participants in their natural environments. The focus on momentary phenomena and immediate reporting is expected to minimize reliance on recall and attendant biases. For example, participants are asked how they are feeling now or how they have felt for the last few minutes rather than being asked to summarize their feelings over hours or days. The implementation of repeated assessments in the participant's natural environment increases the ecological validity of the assessment. Ideally, the assessments constitute a representative sample of the participant's state or behavior in the real world. EMA assessments need not be limited to self-report and may include other assessments, such as ambulatory physiological measures (Shiffman & Stone, 1998).

EMA assessments are often implemented through the use of some cueing device, such as a beeper (Johnson & Larson, 1982), a wristwatch with a programmed alarm (Litt, Cooney, & Morse, 1998), or a palmtop computer (Shiffman et al., 1994). Most often the assessments are scheduled at random intervals to avoid any bias in the sampling of moments. Variants of this approach involve completion of diary records at regular (rather than random) intervals or completion of assessments whenever a target event occurs (e.g., food diaries, self-monitoring). Regardless, EMA methods characteristically involve the repeated assessment of participants' momentary states in their natural environments, which are typically regarded as a representative sample of the participants' condition in general or in particular circumstances (e.g., during meals). The assessments can be aggregated into an accurate and reliable summary of a person's typical condition (i.e., average mood, blood pressure, etc.).

The disadvantage of this methodology is the significant burden imposed on both participants and investigators (Shiffman & Stone, 1998). For example, completion of repeated assessments in the midst of a busy day imposes a considerable burden on participants. This also increases the chances that the sample of persons is biased (because only some people are willing or able to complete the study) or that the sample of assessed moments is biased (because people

are only willing to complete assessments at certain times). For the investigator, these methods may involve large investments in time and funds because of the technology and time intensive data management. Employing palmtop computers (to prompt participants, to present assessments, and to store the data) is a growing trend in EMA research (see Shiffman et al., 1994; Stone, Shiffman, & deVries, in press). In many cases, these disadvantages are outweighed by EMA's distinct advantages (Shiffman & Stone, 1998). Compared with other forms of self-report, EMA data are less subject to biases introduced by recall and retrieval processes. Of course, EMA data are still subject to other sources of biases (e.g., deliberate distortion) that plague all self-report data. For example, there is research showing that certain groups are prone to underreporting food intake such as "diet-resistant" individuals, obese "small eaters," obese adults, younger Caucasian females, respondents with less education, and "restrained eaters" (see Lowe, Kopyt, & Buchwald, 1996, for a complete review). In addition, because real-world occasions are sampled, EMA data may be more ecologically valid, especially in comparison to laboratory data. Finally, EMA methods are uniquely suited to exploring dynamic relationships between variables that interact over time. By studying individuals repeatedly and intensively over time, EMA methods allow the investigator to explore acute or transient (but often important) effects that cannot be discerned in summary data.

A review of the literature reflects a growing trend in the use of EMA methods. EMA methods are being applied to the study of stress and cardiovascular functioning (Kamarck, Shiffman, Smithline, Goodie, & Jong, 1998; Guyll & Contrada, 1998), chronic pain (Affleck et al., 1998; Stone, Broderick, Porter, & Kaell, 1997), cigarette smoking (Shiffman, Paty, Gnys, Kassel, & Hickcox, 1996), stress and coping (Stone, Schwartz, et al., in press), asthma (Smyth, Soefer, Hurewitz, & Stone, 1997), hormones (Okenfels et al., 1995; Smyth, Ockenfels, et al., 1997), and psychiatric disorders (Delespaul, 1995). There are many other areas of health psychology where EMA methods could make a contribution, including eating behavior.

Problems with Self-Monitoring Food Intake

Handheld computers offer a variety of capabilities that make them optimal for nutrition assessment (Jerome, Behar, & Dobbs, 1995; Jerome, Frederiksen, Fredericksen, 1991; Orta & Reinarts, 1994) and are increasingly being utilized in this area both for weight management and diabetes management. It is possible that such technology offers a relatively new methodology that may improve the accuracy of food intake assessment, thus remedying some of the methodological limitations of diary data. This methodology addresses several limitations of existing methods including provision of food lists, relative ease, coding, and novelty. As part of the current study, the utility of palmtop computers in the assessment of food intake was evaluated among overweight women participating in a behavioral weight loss program and among normal-weight age- and ethnicity-matched women (See Sbrocco, Stone, Nedegaard, Lewis, Patel, & Gallant, 1997). Reported food intake was accurate for both the obese and normal weight participants when compared to their expected needs based on calculated energy equations from the American Dietetic Association (1990; which takes into account body weight and activity level), thus providing preliminary evidence that the estimated energy intake from the computer diaries provides accurate data for both normal and obese participants. These energy equations do not allow for an estimation of meal sizes or frequencies, only for daily total caloric intake.

Purpose of Present Study

Currently, without knowing the accuracy of estimates made from retrospective self-report, we don't know what the implications are for the diagnostic process. The purpose of the present study was to address the accuracy of these estimates by comparing retrospective reports from individual questions on the Eating Disorder Examination (EDE, Version 12.0D; Fairburn & Cooper, 1993), a widely used semi-structured clinical interview for diagnosing eating disorders, with self-monitoring data. Retrospective self-report data from the EDE (frequency counts and

reports of cognitive-affective states) were compared with moment-by-moment data recorded in handheld computerized eating diaries by obese and normal-weight women.

Five EDE parameters (meal patterns, restraint, avoidance of eating, types of overeating, and reaction to prescribed weighing) were examined. These scores were compared against self-monitored eating behavior data recorded on palmtop computers. The key questions from the EDE were operationalized using 2 weeks of food diary data as described below.

METHOD

Participants

Participants were 24 obese women and 16 normal weight women, between the ages of 24 - 59. Ideal weight was defined according to the Metropolitan Life Insurance Company Tables (1983) for a medium framed individual. According to the tables, 130% of ideal weight is the cut-off for "obesity" and 160% is considered the upper cut-off for "moderate obesity." In addition, 85% of ideal weight is the upper cut-off for being considered "underweight" and 120% is the cut-off for being "overweight." Therefore, for the purposes of this study, obese participants were 130-160% of ideal weight and normal weight participants were 90-110% of ideal weight. Obese participants were recruited through community advertisement to participate in a weight management study. Normal weight participants were recruited through advertisement and from university staff to participate in a study examining eating behavior of normal weight women. Normal weight participants were paid for their participation. Seventeen obese and 15 normal weight participants provided useable diary data for this study. Data collected from the remaining participants was incomplete and therefore could not be used. Participant characteristics for the current study are shown in Table 7.

Measures

Eating Disorder Examination (EDE)

Participants were administered the Eating Disorder Examination (EDE 12.D; Fairburn & Cooper, 1993). EDEs were conducted by 2 clinical psychology graduate students trained in administration of the EDE and familiar with eating disorders and obesity.

Eating Records

Dietary intake was assessed using the Psion 3.0A palmtop computer (Psion PLC, 1994). The Psion 3.0A contains a 1 megabyte internal disk and several built-in applications (e.g., word processor, agenda). The Psion 3.0A is user-friendly and programs are displayed as icons. Participants highlight icons and press enter to begin a program. Dietary intake was recorded using Comcard COMPUTE-A-DIET Nutrient Balance System software nutrition assessment program (Compute-A-Diet Nutrient Balance System, 1993). The program contains almost 4000 foods from the United States Department of Agriculture (USDA) Data Base. The database entries are listed in alphabetical order and foods can be located by searching alphabetically or by food group type. New food items with their associated nutrient values can be added to the database enabling individuals to expand the existing database to include brand names or personal recipes. Entries are time and date stamped. Participants weighed all foods in grams or ounces using portable scales. Participants also recorded type of meal (breakfast, lunch, dinner, or snack) and situational parameters (location the meal was eaten, who the participant ate with, whether or not the participant overate, how tasty the meal was, and whether or not the participant was upset). Type of meal and situational parameters were recorded using the WEIGHT program, a software program developed for the study.

Procedure

Approval from the university Institutional Review Board was obtained and all participants consented to participate. Obese and normal weight participants met separately in small groups of 5 – 10 at the beginning and end of the 2-week period. They were weighed at each meeting. At the first meeting participants were instructed in use of the Psion palmtop computer. Participants entered practice meals which were evaluated by study research staff before participants left. Participants were given further individual instruction and evaluation as needed and provided a “hotline” number which they could call during business hours for help. Participants also completed self-report measures and scheduled an appointment to complete the Eating Disorders Examination and to assess body composition during the two-week period. Participants met as a group for the second session two weeks later and turned in their diaries. All participants were provided with written feedback from their diaries including reported daily caloric intake and macronutrient intake.

Data Reduction and Analyses

Five variables were calculated from the eating records in order for comparisons to be made with key EDE responses:

Frequency of meals (breakfast, mid-morning snack, lunch, mid-afternoon snack, evening meal, and evening snack). This allowed for a comparison to be made with the pattern of eating from the EDE.

Daily caloric intake. This allowed for comparisons to be made with restraint over eating, days of overeating, and reaction to prescribed weighing from the EDE.

Frequency of days with 8 hours between meals. This allowed for a comparison to be made with avoidance of eating from the EDE.

Caloric intake by meal. This allowed for a comparison to be made with episodes of overeating from the EDE.

Frequency of meals self-reported as “overeaten.” This allowed for an additional comparison to be made with episodes of overeating from the EDE.

Five domains from the EDE were selected for comparisons with the variables calculated from the food diaries:

1. Pattern of Eating

Pattern of eating is assessed using the EDE by asking individuals how often they consume 6 meals types: breakfast, mid-morning snack, lunch, mid-afternoon snack, evening meal, evening snack, and nocturnal snack. These frequency ratings are made on a 7 point scale that is anchored by descriptors at 4 points (0 = not eaten; 2 = eaten on less than half the days; 4 = eaten more than half the days; 6 = eaten every day). Using the computerized diaries, individuals categorized meals as a snack, breakfast, lunch, or dinner, and also recorded the time of day. In order to make comparisons between the EDE-generated frequencies and the diary frequencies, diary snacks had to be categorized into mid-morning, mid-afternoon, or evening. In addition, the evening and nocturnal snack categories were collapsed because it was not clear from diary entries whether the individual had been sleeping and awoke to consume the snack. For each meal, the percentage of days the type of meal was eaten was calculated from the diaries. This percentage data was then transformed into the EDE categorization (0 = not eaten to 6 = eaten every day). The EDE frequency ratings for the six meals were compared with the frequency and percentage of meals entered in the computer diaries.

2. Restraint Over Eating

Restraint over eating is assessed using the EDE by asking individuals to rate the number of days they have consciously attempted to restrict what they eat. Individuals rate the number of days on a 7 point scale with four anchors (0 = no restraint; 2 = restraint on less than half the days; 4 = restraint on more than half the days; 6 = restraint every day). Because a direct measure of restraint was not available from the diaries—that is, participants did not record whether they were

restricting their intake—restraint was operationally defined as calorie overrestriction. In reality, it is unknown why subjects recorded days with low calorie intake. Overrestriction was conservatively operationalized as daily intake at least two standard deviations below mean intake. This was 1362 kcal for obese and 761 kcal for normal weight women. The percentage of days of overrestriction was calculated from the diaries. This percentage data was then transformed into the EDE categorization (0 = no restraint to 6 = restraint every day). The EDE frequency ratings for restraint were compared with the frequency and percentage of overrestricted days entered in the computer diaries.

3. Avoidance of Eating

Avoidance of eating is examined as another component of dietary restraint. Avoidance is assessed by asking individuals whether they have gone for periods of 8 or more waking hours without eating. Individuals rate the number of days during which there has been at least 8 hours of abstinence on a 7 point scale with four anchors (0 = no avoidance; 2 = avoidance on less than half the days; 4 = avoidance on more than half the days; 6 = avoidance every day). Avoidance of eating was assessed from the diaries by calculating the number of 8-hour breaks during waking hours (8 a.m. to 11 p.m.). Of course, there are other possible reasons subjects did not eat during an 8-hour period, such as being too busy or unavailability of food. The percentage of days with avoidance was calculated from the diaries. This percentage data was then transformed into the EDE categorization (0 = no avoidance to 6 = avoidance every day). The EDE frequency ratings for avoidance were compared with the frequency and percentage of days with avoidance entered in the computer diaries.

4. Episodes of Overeating

Overeating is assessed using the EDE by asking individuals to recall the number of episodes and days they felt they ate too much. Four different types of overeating are distinguished on the EDE: objective bulimic episodes, objective overeating, subjective bulimic

episodes, and subjective overeating. The distinctions among these four categories of overeating are based on the presence or absence of the basic criteria for a binge episode: loss of control and the consumption of what most people would regard as a large amount of food. Participants did not record loss of control over eating in the computerized diaries but they did report their perceptions of whether they overate or not for each meal. Consequently, overeating could not be distinguished from binge eating. The following frequencies from the EDE were used to assess overeating: number of overeating episodes and number of days during which overeating occurred. This information was compared with data from the diaries on: number of incidences of self-reported overeating, number of high calorie meals, and number of high calorie days. Overeating episodes were assessed by computing the frequency and number of meals the individuals reported they overate. Objective overeating at meals were operationalized as those exceeding 1000 kcal for both groups, because Fairburn (1987) recommended using a cut-off of an average consumption per binge of 1000 kcal for a diagnosis of bulimia nervosa. Even though binge eating was not specifically defined from the computerized diary data, it was felt that this would still be an interesting comparison. Objective overeating was also operationalized as high calorie days where intake exceeded twice the standard deviation above the mean (3560 kcal for obese; 3297 kcal for normals).

5. Reaction to Prescribed Weighing

Reaction to prescribed weighing is assessed using the EDE by asking individuals how they would feel if they were asked to weigh themselves once each week for the next 4 weeks.

The interviewer rates the strength of the respondent's reaction on a 7 point scale with four anchors (0 = no reaction; 2 = slight reaction; 4 = moderate reaction; 6 = marked reaction).

Because a direct measure of reaction was not available from the diaries—that is, participants did not record their reaction to being weighed—reaction to prescribed weighing was operationally defined as consuming less kcal during the four days prior to being weighed at the end of the two

week baseline period than during the first ten days. It was hypothesized that participants who were reactive to being weighed would restrict their daily caloric intake at the end of the baseline period in preparation for being weighed. In reality, there may have been other reasons subjects ate less at the end of the baseline period. Mean daily caloric intake for the last four days was subtracted from the first ten days for each participant. This difference was then transformed into the EDE categorization (0 = no reaction to 6 = marked reaction) using a 7 point scale (0 = last four days not less than first ten days; 1 = difference of 167 kcal; 2 = difference of 333 kcal; 3 = difference of 500 kcal; 4 = difference of 667 kcal; 5 = difference of 833 kcal; 6 = difference of 1000 kcal). The EDE ratings for reaction to prescribed weighing were compared with the kcal differences and ratings calculated from the computer diaries.

Diary data was summarized using the Self-Monitoring Analysis System (Version 4.0) (SMAS; Schlundt, 1993). SMAS is a computer software system designed primarily to be used to aid in behavioral assessment and behavioral analysis. Specifically, the primary aim in the development of SMAS was to design a software system that would enhance the ability to use self-monitoring diaries as a behavioral assessment tool. For each of the five domains, the diary data and their EDE counterparts were correlated to determine the degree of similarity in the ratings obtained by each measure. These analyses were made for obese participants, normal-weight participants, and all participants combined. Previous research suggests that responses of obese and normal weight women may be different among these domains. The distributions were then compared between EDE and diary variables to examine for agreement. Because most of the variables were nonnormally distributed, nonparametric tests were used. Kendall's tau-b was used for correlations between the EDE and diary variables and Wilcoxon's Matched-Pairs Signed-Ranks Test was used to compare distributions between the EDE and diary variables. Results from the Kendall's tau-b and Wilcoxon's Matched-Pairs Signed-Ranks Test between the EDE and diary variables for the obese subjects, normal-weight participants, and all subjects combined are presented for each domain in Tables 8 – 14. In addition, the frequency distributions for the EDE

and diary variables are displayed in Figures 3 – 7, allowing for a visual comparison of the distributions. For all statistical analyses, the alpha level was established a priori at .05 for statistical significance. However, because of the small sample size in this study, it was decided that *p*-values between .05 and .10 would be reported as statistical “trends.” It is expected that future studies using a larger number of subjects would find statistical significance in analyses reported as “trends” in the present study.

RESULTS

1. Pattern of Eating. Presented in Table 8 are comparisons for the obese, normal, and combined samples between the EDE and the computerized diary, by meal. The mean frequency ratings (0 – 6) for each meal type from the EDE and the computerized diaries, as well as the percentage of days from the diaries on which meals were eaten, are shown. Results from Kendall’s tau-b and Wilcoxon’s Matched-Pairs Signed-Ranks Test comparison of the frequencies (0 – 6) for each meal are listed. Kendall’s tau-b correlations between the EDE and the diary for dinner ($\tau = .50, p < .05$) and evening snacks ($\tau = .38, p < .05$) were statistically significant for all the participants combined. Breakfast ($\tau = .48, p < .05$) and dinner ($\tau = .50, p < .05$) were significantly correlated for the normal-weight participants. Lunch was the meal least correlated for both the obese ($\tau = .02, p > .10$) and the normals ($\tau = .13, p > .10$). Overall, the responses of the normal weight individuals were better correlated than the obese participants’ responses. A Wilcoxon Matched-Pairs Signed-Ranks Test was applied to the ranked data for meals reported from the EDE and diaries. For all the subjects combined, the distributions were statistically different for morning snacks ($T = -2.37, p < .05$) and dinner ($T = -2.28, p < .05$). Superscripts in Table 8 denote whether ratings were higher from the EDE or the diaries. The distributions between the EDE and the diaries for breakfast ($T = -2.09, p < .05$) and lunch ($T = -2.33, p < .05$) among the obese sample were statistically different and the distributions for

morning snacks ($T = -2.08, p < .05$) and dinner ($T = -2.14, p < .05$) were statistically different among the normal-weight subjects.

Despite the overall similarity in mean frequency estimates from the EDE and diaries and the significant correlation between the EDE and diaries for some meals, there are large discrepancies in distributions between the EDE and the computerized diaries. These discrepancies are illustrated in Figure 3 and Figure 4, by meal. The EDE estimates are depicted on the vertical axes and the corresponding diary reports are shown on the horizontal axes. As can be seen from Figure 3, most of the obese participants entered breakfast in the diaries every day but on the EDE they reported a greater range of frequencies (from 1 to 6). The majority of the obese's EDE reports for midmorning snacks were anchored at 0 and 6 (never or always) but the diary reports were predominantly scores of 0 and 1. The obese participants reported on the EDE that they almost always ate lunch but the diaries reflected a range of frequencies of lunch. The linear regression fit lines depict the poor matches between the EDE and diary for mid-morning snacks and lunch. The lines are horizontal rather than diagonal, as they would be with a perfect fit. Although the obese participants reported a broad range of frequencies for mid-afternoon snacks on both the EDE and diaries, the match was very poor. All but 1 of the obese participants reported on the EDE that they ate dinner every day, but a range of frequencies was actually recorded in the diaries. Finally, as depicted in Figure 3, most of the obese participants reported evening snack EDE scores of 0 – 3 but they recorded a much broader range of frequencies (1 – 6) in the diaries. This is evidenced in the trend for the distributions to differ in Table 8 while the estimates were similar enough to approach statistical significance for tau-b correlations.

Figure 4 demonstrates that the normal-weight participants reported a similar frequency range for meals on the EDE and diaries. Except for lunch, there was at least a statistical trend for all meals between the two instruments to be correlated.

2. Restraint over Eating. Presented in Table 9 are comparisons for the obese, normal, and combined samples between the EDE and the computerized diary for restraint over eating. The mean ratings (0 – 6) from both the EDE and computerized diaries for days of restricted eating, as well as the percentage of days from the diaries on which caloric intake was less than 2 standard deviations below the mean, are shown. Results from Kendall's tau-b and Wilcoxon's Matched-Pairs Signed-Ranks Test comparison of the restraint ratings (0 – 6) between the EDE and diaries are listed. Kendall's tau-b correlations between the EDE and diaries were not statistically significant for all the participants combined ($\tau = .18, p > .10$), the obese subjects ($\tau = .02, p > .10$), and the normal-weight participants ($\tau = .11, p > .10$). A Wilcoxon's Matched-Pairs Signed-Ranks Test was applied to the ranked data for reported restraint over eating from the EDE and calorie-restricted days from the diaries. For all the subjects combined the distributions were not statistically different ($T = -0.64, p > .10$). However, there was a trend for the obese' ($T = -1.79, p < .10$) and the normals' ($T = -1.65, p < .10$) distributions to be statistically different. Superscripts in Table 8 denote whether ratings were higher from the EDE or the diaries. In summary, as depicted in Table 9, there was no significant association between reported restraint over eating on the EDE and percentage of low-calorie days from the eating diaries. As further shown in Figure 5, all but one of the normal weight women and 5 of the obese participants indicated no restraint on the EDE (score = 0). However, based on the diary data, all but 2 obese individuals and 9 normal weight subjects did have low calorie days. The limited range of responses on the EDE by normals prevented the calculation of a linear regression fit line. It is important to note that there is no way of knowing if low calorie days reflect conscious restriction and they may, in fact, be due to other factors such as illness or being too busy to eat.

3. Avoidance of Eating. Presented in Table 10 are comparisons for the obese, normal, and combined samples between the EDE and the computerized diaries for avoidance of eating. The mean ratings (0 – 6) from both the EDE and computerized diaries for days during which at least 8

hours passed without eating, as well as the percentage of days from the diaries on which there were periods of 8 continuous hours without eating, are shown. Results from Kendall's tau-b and Wilcoxon's Matched-Pairs Signed-Ranks Test comparison of the avoidance ratings (0 – 6) between the EDE and diaries are listed. As shown in Table 10, reported avoidance of eating from the EDE and percentage of days with 8 hours between meals was non-significantly correlated ($\tau = .27$, $p < .10$, for the combined sample and $\tau = .42$, $p < .10$, for the obese). For the normal and obese participants, respectively, 23% and 19% of the days had at least one 8-hour break. However, as depicted in Figure 6, all but one of the obese and all of the normal participants reported on the EDE they never avoided eating. This fact was reflected in the significant differences between distributions in Table 10. Wilcoxon's Matched-Pairs z scores were $T = -4.0$, $p < .05$, for the combined sample; $T = -2.4$, $p < .05$, for the obese; and $T = -3.25$, $p < .05$ for the normals. This also explains why neither Kendall's tau-b nor a linear regression fit line could be calculated for the normal weight participants.

4. Overeating. Presented in Table 11 are comparisons for the obese, normal, and combined samples between the EDE and the computerized diaries for days of overeating. The mean number of reported days of overeating from the EDE, the mean number of days on which more than 2 standard deviations above the mean daily caloric intake from the computerized diaries, and the mean number of kcals consumed per day are shown. The mean kcal/day was 2462 ($SD = 550$) for the obese and 2029 ($SD = 634$) for the normal weight participants. Results from Kendall's tau-b and Wilcoxon's Matched-Pairs Signed-Ranks Test comparison of the days of overeating between the EDE and diaries are listed. There was poor correspondence between the EDE and diaries ($\tau = .11$, $p > .10$, for all subjects and $\tau = .16$, $p > .10$ for the obese). A Wilcoxon's Matched-Pairs Signed-Ranks Test demonstrated a significant difference between distributions for the normal-weight subjects ($T = -2.82$, $p < .05$) and a trend for a significant difference for the obese subjects ($T = -1.53$, $p < .10$). Presented in Table 12 are comparisons for

the obese, normal, and combined samples between the EDE and the computerized diaries for episodes of overeating. The mean number of reported episodes of overeating from the EDE, the mean number of meals recorded on the diaries in excess of 1000 kcals, and the mean number of meals self-reported on the diaries as “overeaten” are shown. Results from Kendall’s tau-b and Wilcoxon’s Matched-Pairs Signed-Ranks Test comparison of the episodes of overeating between the EDE and diaries are listed. There was poor correspondence between the EDE and computerized diaries for high calorie meals ($\tau = .16$, $p > .05$, for all subjects and $\tau = .23$, $p > .05$ for obese subjects). There was also poor correspondence between the EDE and computerized diaries for meals recorded as “overeaten” ($\tau = .32$, $p > .05$, for all subjects and $\tau = .08$, $p > .05$, for obese subjects). A Wilcoxon’s Matched-Pairs Signed-Ranks Test demonstrated statistically significant differences between the distributions. Approximately 20% of all meals were greater than 1000 kcals. As depicted in Table 13, self-reported overeating occurred fairly often: 45% ($SD = 34$) of the obese’s meals and 26% ($SD = 24$) of the normals’ meals. The mean kcal/meal was 553 ($SD = 128$) for the obese and 690 ($SD = 358$) for the normal weight participants. The obese participants averaged 766 kcal ($SD = 500$) when they reportedly overate and 388 kcal ($SD = 218$) when they did not. Normal weight participants averaged 624 kcal ($SD = 408$) when they reportedly overate and 552 kcal ($SD = 218$) when they did not.

5. Reaction to Prescribed Weighing. Presented in Table 14 are comparisons for the obese, normal, and combined samples between the EDE and the computerized diary for reactions to prescribed weighing. The mean ratings (0 – 6) of reaction to prescribed weighing from the EDE, the mean ratings (0 – 6) of differences in daily caloric intake during the last 4 days of the baseline compared to the first 10 days from the computerized diaries, and the actual differences in daily caloric intake during the last 4 days of the baseline compared to the first 10 days from the diaries are shown. Results from Kendall’s tau-b and Wilcoxon’s Matched-Pairs Signed-Ranks Test comparison of the reaction to prescribed weighing ratings (0 – 6) between the EDE and diaries

are listed. As can be seen in Table 14, reported reaction to prescribed weighing was not related to the difference between early and later daily caloric intake. There was poor correspondence between the EDE and diary variables ($\tau = -.06$, $p > .10$, for all subjects; $\tau = -.13$, $p > .10$, for the obese; and $\tau = -.07$, $p > .10$, for the normals). There were also statistically significant or trend differences between the distributions ($\underline{T} = -2.93$, $p < .05$, for all subjects; $\underline{T} = -1.48$, $p < .10$, for the obese; and $\underline{T} = -2.84$, $p < .05$, for the normals). According to Figure 7, most participants reported no reaction to prescribed weighing on the EDE, but most participants ate fewer kcals the four days prior to being weighed at the end of the baseline period. As a group, the participants averaged 157 kcals/day less than over the first ten days of the baseline period.

DISCUSSION

These results suggest some lack of correspondence between the diary data and the EDE for a frequency count of most meal types and for overeating days and episodes, as well as for most self-reported cognitive-affective states. Many responses on the EDE appeared anchored at either end, reflecting endorsements of daily or never. However, moment-by-moment recording in the eating diary reflected a range of responses. Correspondence was better for certain meals (i.e., normals' breakfasts and dinners). It should be reemphasized that some of the variables operationalized from the eating diaries are based on imperfect interpretations in order for comparisons to be made with the EDE. For example, it is unclear if participants went more than 8 waking hours without eating by choice or because of environmental circumstances. Likewise, it is unknown if participants ate alone because they did not want others to see them eat.

The purpose of this study was to evaluate the validity of retrospective information reported on the EDE. The answer to this question probably depends on the intended use of the instrument. Because there was generally poor correspondence between the EDE and ecological momentary assessment of specific details (e.g., frequency counts), the instrument probably is a poor choice to use to obtain quantitative data. Because the DSM-IV criteria for Bulimia Nervosa

and Binge Eating Disorder require quantitative data to make these diagnoses (i.e., bingeing and purging at least twice a week), the intended use of version 12.0D of the EDE for assessing DSM-IV eating disorders is called into question. In fact, the results of this study, and others like it, call into question the validity of using retrospective frequency counts for making any DSM-IV diagnosis. Another questionable use of the EDE would be for research studies in which quantitative data or specific details are sought for evaluation. When quantitative data or specific details are required, a measure that validly taps into episodic memory must be employed. Ecological momentary assessment measures, such as a moment-by-moment eating diary, is clearly a better choice than retrospective self-report instruments, such as the EDE.

The EDE appears to be a useful instrument, however, for collecting qualitative information. Many researchers believe that disordered eating exists on a continuum (e.g., Lowe et al., 1996; Schlundt & Johnson, 1990). If that is the case, then precise frequency counts and specific details become less important in determining how "pathological" a person is than in trying to fit a person into a category. In a previously reported study (Stone, Nedegaard, & Sbrocco, 1997), five of the obese participants from the current study reported on the EDE that they had binge eaten. The obese non-bingers fell between the normal weights and obese bingers on all subscales of the EDE. For all subscales the non-bingeing obese scored significantly lower than the obese bingers and significantly higher than the normals. For the self-report measures, the same pattern was observed for the Binge Scale Questionnaire and two of the three subscales of the EDI-2 relating to eating-specific pathology (Bulimia, Drive for Thinness) and the Hunger subscale of the Eating Inventory. Obese binge eaters and non binge eaters did not differ from each other but scored higher than normals on the Body Dissatisfaction Subscale of the EDI-2 and the Disinhibition subscale of the Eating Inventory. Lastly, normals and obese non-bingers did not differ on most of the EDI-2 subscales designed to assess more general pathology associated with eating disorders, while obese binge eaters did. Normals and obese non-bingers also did not differ on the Restraint subscale of the EI, despite differences on the Restraint subscale of the

EDE. The primary purpose of the Stone, Nedegaard, and Sbrocco (1997) study was to examine whether obese bingers, obese non-bingers, and normal weight women fell on a continuum when measuring eating related pathology. This appeared to be the case for most measures of eating related pathology. The EDE proved to be an effective instrument for assessing the degree of disordered eating pathology among obese women, as its subscales demonstrated concurrent validity with other disordered eating instruments. In this case, exact episodic memory was not necessary as more generic memory sufficed.

The results of this study are not surprising, in light of Menon's (1993) proposed model for the storage and retrieval of information about frequent behavior. The behaviors measured in the current study (frequency of meals, frequency of restrained eating, episodes of avoidance, episodes of overeating, and reaction to prescribed weighing) would best be described as irregular-similar behaviors, given that they occur on an irregular basis but are similar when they do occur. According to Menon (1993), recalling irregular-similar behaviors requires the greatest cognitive effort because it involves the most complex process of memory retrieval. There is no readily accessible rate of occurrence stored in the semantic store because of the irregularity of the behavior, and the respondent will need to use more cognitive effort in order to estimate the behavioral frequency using some strategy other than a general rate-based one. This may explain why many responses on the EDE appeared anchored at either end, reflecting endorsements of daily or never. Many respondents may not have taken the cognitive effort to devise strategies to accurately recall the behaviors in question. In essence, they may have taken "the easy way out" by endorsing one extreme or the other.

These results have specific implications for retrospective self-report methodology and general implications for research on the cognitive processes that subserve recall data. Asking people to recall their dietary intake and cognitive-affective states for any length of time, or for arbitrarily remote periods, appears impractical if specific, detailed reports are desired. Still, although individuals' reports are certainly not comprehensive for specific periods, and contain

many intrusions, they probably reflect, to some extent, the typical diets of individuals. Thus, participants' reports are more like diet histories—descriptions of their typical diets—than like dietary recalls. If dietary reports are based substantially on generic memory, perhaps generic memory is what researchers and clinicians should ask about. Perhaps, when the reference period of interest exceeds the previous few hours, it is futile to ask individuals to report their intake and to expect report of specific memories.

The news is not all bad. Cognitive psychologists and survey researchers are studying cognitive processes and devising and evaluating methods to improve recall and reduce reporting errors and bias (Friedenreich, 1994). These researchers have found that question comprehension can be improved by increasing the question length, by providing more instructions, by using simpler wording, and by changing the question order to be more compatible with autobiographical memory (asking first “why,” then “how,” and, finally, “when”) (Means et al., 1991; Jobe, Tourangeau, & Smith, 1994; Jobe & Mingay, 1989). A vital step in improving question comprehension has been to test the questionnaire first in a laboratory setting and then in a field pre-test (Bercini, 1992).

To improve information retrieval, cognitive interviewing techniques have been applied to food-consumption recall (Fisher, Quigley & Kathryn, 1992). The cognitive interview is based on principles of memory retrieval, cognition, and communication, as gathered from laboratory research in cognitive psychology and from extensive analysis of tape-recorded interviews conducted in the field (Fisher, Geiselman, & Raymond, 1987). Although the cognitive interview was developed initially to enhance witness recollection in criminal investigations, with minor modifications it has been shown to be applicable to other types of investigative interviewing, since it is based primarily on general principles of cognition. In an experiment by Fisher, Quigley, and Kathryn (1992), more than twice as many foods were recalled with the cognitive interview than with the no-instruction interview, with no more intrusion errors.

The theoretical framework of the cognitive interview, as applied to eliciting food-consumption histories, is based on five general principles of cognition and memory retrieval: context reinstatement, focused retrieval, extensive retrieval, varied retrieval, and multiple representations (Fisher, Quigley & Kathyrn, 1992). Below is provided a simple step-by-step description of each principle.

The principle of context reinstatement suggests that an event will be better recalled if the rememberer is in the same psychological environment as when the event occurred originally (Tulving and Thomson, 1973). Thus, the interviewer encourages the respondent to think about the environmental context (e.g., dining-room arrangement, lighting conditions) and the relevant psychological context (e.g., why the respondent selected specific foods) at the time of the original meal.

The principle of focused retrieval is based on the concept that memory retrieval, especially for details, requires mental concentration (e.g., Kahneman, 1973). Any distractions from this mental concentrations, whether physical (e.g., extraneous noise) or psychological (e.g., interrupting the respondent's narration to ask a question), will disrupt memory retrieval.

The principle of extensive retrieval states that the more retrieval attempts one makes, the more successful recall will be (Roediger and Payne, 1982). In practice this principle is fulfilled by encouraging respondents to search through memory even if they indicate that they have recalled as much as possible. It is important to note that when the respondent is encouraged to make additional searches through memory, the interviewer cannot simply ask the same question as posed originally. Such an approach often leads the respondent to indicate, "I already told you, I don't know." The interviewer must vary the form of the question, at least superficially, so that the respondent is induced to make another search through memory.

The concept of varied retrieval is based on the notion that memories not activated by one retrieval probe may be accessed with another probe (Anderson and Pichert, 1978). Thus, a respondent who cannot recall a fact when asked a direct question (e.g., "Did you eat any

vegetables?”) may provide the answer when asked a different question (“What foods did you select last?”). In general changing the dimension of the question (e.g., from categorical to temporal) is likely to elicit new information.

The final principle, multiple representations, suggests that the event to be remembered is stored in several forms (Fisher and Chandler, 1984). For example, the foods eaten are stored in several “mental images.” There may be one image of the food while on the plate, another image of the food being served by the waiter, a third image of the food being carried away after the meal, and so on. Each of these images should be evoked and probed separately, exploiting all of the contents before probing other images (Fisher and Quigley, 1988). In order to be aware of the respondent’s multiple representations of the meal, at the beginning of the interview, the interviewer should encourage the respondent to provide an open-ended description of all of the events related to the meal in question. During this initial narration the interviewer should note the various mental representations the respondent has of the meal and probe them later.

Clearly cognitive interviewing techniques, such as those described in the Fisher, Quigley, and Kathryn (1992) study, have the potential to enhance the recall of information queried in the EDE. Future studies should explore this possibility.

Researchers have yet to systematically incorporate methods for improving recall accuracy and reproducibility into the data collection procedures used in retrospective studies (Friedenreich, 1994). Research should be focused on studying the predictors of recall ability to devise methods that reduce recall errors and bias. If these predictors are demographic characteristics associated with particular subgroups of the population with specific recall problems, then different interviewing techniques customized to each subgroup’s needs could be developed. Cognitive research is also necessary to ascertain the specific limits of our memory for recent food intake. For example, at what point do we forget any reasonable amount of detail about actual food consumption of a given day? At what point is it more beneficial to ask about usual food intake or

generic representations of food intake? What other social and cognitive factors affect food recall? More refined measurement of social influence is also needed to better assess this construct.

A large component of recall errors and bias will likely be attributable to aspects of the questionnaire design and interviewing methods. Adopting the cognitive interviewing techniques, questionnaire design, and field pre-testing strategies, as developed by cognitive psychologists and survey researchers, should result in improved data quality in retrospective self-report. All of these methods will require more preparation, testing, and interviewing time. The clear benefits, however, are the improved validity and reproducibility of retrospective data, which should be a principal objective of any researcher or clinician relying on data recalled from the distant past.

Eating diaries normally provide more extensive data on eating behavior than can be obtained from interview or questionnaires. In addition, these diaries provide for momentary assessment, where the participant can report eating immediately after the episode rather than having to rely on extensive recall. While the accuracy of diary food-intake estimates has received extensive criticism, this appears to be less likely to influence the findings of this study. In a previous examination of this method the overall accuracy of food-intake appeared good (Sbrocco, et al., 1997). Therefore, we do not feel that the questionable accuracy of diary data found in many studies is a major limitation of this investigation. However, it is quite possible that meal patterning may be changed by the task demand. This study has several other limitations. First, the eating diaries assessed food intake over only 2 weeks rather than 4 as the EDE does. While not providing a one-to-one match of the retrospective period, given that this study suggests individuals underestimated responses on the EDE, adding two weeks would probably not change the major findings. Clearly, the small sample size of this study limits the findings as does the absence of disordered criterion groups and male participants. One might surmise that more pathology would be associated with more bias during recall; however, the opposite could also be true in that more obsessiveness might lead to greater accuracy. Further data is needed before firm conclusions can be drawn.

In conclusion, because there was generally poor correspondence between the EDE and ecological momentary assessment of specific details (e.g., frequency counts), the instrument probably is a poor choice to use to obtain quantitative data. When quantitative data or specific details are required, a measure that validly taps into episodic memory should be employed. Ecological momentary assessment measures, such as a moment-by-moment eating diary, is clearly a better choice than retrospective self-report instruments, such as the EDE. Unfortunately, food diaries are not always available to clinicians to help make evaluations. The good news is that cognitive psychologists and survey researchers are studying cognitive processes and devising and evaluating methods to improve recall and reduce reporting errors and bias (Friedenreich, 1994). In particular cognitive interviewing techniques have been applied to food-consumption recall in order to improve information retrieval (Fisher, Quigley & Kathryn, 1992). Cognitive techniques such as these show tremendous potential for improving the accuracy of retrospective self-reported details on instruments such as the EDE. Future studies should compare the results of the EDE with and without the use of various cognitive techniques, using ecological momentary assessment as the gold standard. Contributions from the field of cognitive science may one day close the gap between the EDE and food diaries in terms of the accuracy of self-reported eating behavior.

References

- Affleck, G., Tennen, H., Urrows, S., Higgins, P., Abeles, M., Hall, C., Karoly, P., & Newton, C. (1998). Fibromyalgia and women's pursuit of personal goals: A daily process analysis. Health Psychology, 17, 40-47.
- Anderson, R.C., and Pichert, J.W. (1978). Recall of previously unrecalable information following a shift in perspective. Journal of Verbal Learning and Verbal Behavior, 17, 1-12.
- American Psychiatric Association. (1987). Diagnostic and statistical manual of mental disorders (3rd ed., rev.). Washington, D.C.: Author.
- American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (4th ed.). Washington, D.C.: Author.
- Bandini, L.G., Schoeller, D.A., Cyr, H.N., & Dietz, W.H. (1990). Validity of reported energy intake in obese and nonobese adolescents. American Journal of Clinical Nutrition, 52, 421-425.
- Beglin, S.J. (1990). Eating disorders in young adult women. Unpublished doctoral dissertation, Oxford University, Oxford.
- Bercini, D.H. (1992). Pretesting questionnaires in the laboratory: An alternative approach. Journal of Experimental and Analytical Environmental Epidemiology, 2, 241-248.
- Blair, E., & Burton, S. (1987). Cognitive processes used by survey respondents to answer behavioral frequency questions. Journal of Consumer Research, 14, 280-288.
- Beumont, P.J.V., Kopec-Schrader, E.M., Talbot, P., & Touyz, S.W. (1994). Measuring the specific psychopathology of eating disorder patients. Australian and New Zeland Journal of Psychiatry.
- Brownell, K.D. (1994). The LEARN program for weight control (6th ed.). Dallas, TX: American Health.

Bunnell, D.W., Shenker, I.R., Nussbaum, M.P., Jacobson, M.S., & Cooper, P.J. (1990). Subclinical versus formal eating disorders: Differentiating psychological features. International Journal of Eating Disorders, 9, 357-362.

Clinton, D.N., & Glant, R. (1992). The eating disorders spectrum of DSM-III-R. Clinical features and psychosocial concomitants of 86 consecutive cases from a Swedish urban catchment area. Journal of Nervous and Mental Disease, 180, 244-250.

Cofer, C.N. (1967). Does conceptual organization influence the amount retained in immediate free recall? In B. Kleinmuntz (Ed.), Concepts and the structure of memory (pp. 181-214). New York: Wiley.

Compute-A-Diet Nutrient Balance System [Computer Software]. (1993).
Worcestershire, England: COMCARD LTD.

Cooper, P.J., Taylor, M.J., Cooper, Z., & Fairburn, C.G. (1987). The development and validation of the Body Shape Questionnaire. International Journal of Eating Disorders, 6, 485-494.

Cooper, Z., & Fairburn, C. (1987). The Eating Disorder Examination: A semi-structured interview for the assessment of the specific psychopathology of eating disorders. International Journal of Eating Disorders, 6, 1-8.

Cooper, Z., Cooper, P.J., & Fairburn, C.G. (1989). The validity of the Eating Disorder Examination and its subscales. British Journal of Psychiatry, 154, 807-812.

Deese, J. (1961). From the isolated verbal unit to connected discourse. In C. Cofer (ed.), Verbal learning and verbal behavior (pp. 11-31). New York: McGraw-Hill.

Delespaul, P. (1995). Assessing schizophrenia in daily life—the experience sampling method. Maastricht, the Netherlands: Maastricht University Press.

Drougas, H.L., Reed, G., & Hill, J.O. (1992). Comparison of dietary self-reports with energy expenditure measured using a whole-room indirect calorimeter. Journal of the American Dietetic Association, 92, 1073-1077.

Ebbinghaus, H. (1885/1964). Memory: A contribution to experimental psychology. New York: Dover.

Erdelyi, M.H., Finks, J., & Feigin-Pfau, M.B. (1989). The effect of response bias on recall performance, with some observations on processing bias. Journal of Experimental Psychology: General, 118, 245-254.

Fairburn, C.G., Cooper, Z., & Cooper, P.J. (1986). The clinical features and maintenance of bulimia nervosa. In K.D. Brownell & J.P. Foreyt (eds.), Handbook of eating disorders: Physiology, psychology and treatment of obesity, anorexia and bulimia. New York: Basic Books.

Fairburn, C.G., Jones, R., Peveler, R.C., Carr, S., Solomon, R., O'Connor, M.E., & Hope, R.A. (1991). Three psychological treatments for bulimia nervosa: A comparative trial. Archives of General Psychiatry, 48, 463-469.

Fairburn, C.G., & Cooper, Z. (1993). The Eating Disorder Examination (12th Edition). In C.G. Fairburn & G.T. Wilson (Eds.), Binge eating: Nature, assessment, and treatment (pp. 317 - 356). New York: Guilford Press.

Fairburn, C.G., Jones, R., Peveler, R.C., Hope, R.A., & O'Connor, M. (1994). Psychotherapy and bulimia nervosa: The longer-term effects of interpersonal psychotherapy, behaviour therapy and cognitive behaviour therapy. Archives of General Psychiatry.

Fairburn, C.G., Peveler, R.C., Davies, B., Mann, J.I., & Mayou, R.A. (1991). Eating disorders in young adults with insulin dependent diabetes mellitus: A controlled study. British Medical Journal, 303, 17-20.

Fairburn, C.G., Peveler, R.C., Jones, R., Hope, R.A., & Doll, H.A. (1994). Predictors of twelve-month outcome in bulimia nervosa and the influence of attitudes to shape and weight. Journal of Consulting and Clinical Psychology.

Fairburn, C.G., Stein, A.P., & Jones, R. (1992). Eating habits and eating disorders during pregnancy. Psychosomatic Medicine, 54, 665-672.

Fisher, R.P., & Chandler, C.C. (1984). Dissociations between temporally-cued and theme-cued recall. Bulletin of the Psychonomic Society, 22, 395-397.

Fisher, R.P., Geiselman, R.E., & Raymond, D.S. (1987). Critical analysis of police interview techniques. Journal of Police Science and Administration, 15, 177-185.

Fisher, R.P., & Quigley, K.L. (1988). The effect of question sequence on eyewitness recall. Manuscript. Miami, FL: Florida International University.

Fisher, R.P., Quigley, K.L., & Kathyrn, L. (1992). Applying cognitive theory in public health investigations: Enhancing food recall with the cognitive interview. In J.M. Tanur, et al. (eds.), Questions about questions: Inquiries into the cognitive bases of surveys (pp. 154-169). New York: Russell Sage Foundation.

Fricker, J., Baelde, D., Igoïn-Apfelbaum, L., Huet, J.N., & Apfelbaum, M. (1992). Underreporting of food intake in obese "small eaters." Appetite, 19, 273-283.

Friedenreich, C.M. (1994). Improving long-term recall in epidemiologic studies. Epidemiology, 5, 1-4.

Garfinkel, P.E., & Garner, D.M. (1982). Anorexia Nervosa: A multidimensional perspective. New York: Brunner Mazel.

Garner, D.M., & Garfinkel, P.E. (1979). The Eating Attitudes Test: An index of the symptoms of anorexia nervosa. Psychological Medicine, 9, 273-279.

Garner, D.M., Olmstead, M.A., & Polivy, J. (1983). Development and validation of a multidimensional Eating Disorder Inventory for anorexia nervosa and bulimia nervosa. International Journal of Eating Disorders, 2, 15-34.

Garner, D.M., Rockert, W., Davis, R., Garner, M.V., Olmsted, M.P., & Eagle, M. (1993). Comparison between cognitive-behavioral and supportive-expressive therapy for bulimia nervosa. American Journal of Psychiatry, 150, 37-46.

Giles, T., Young, R., & Young, D. (1985). Behavioral treatment of severe bulimia. Behavior Therapy, 16, 393-405.

Gormally, J., Black, S., Daston, S., & Rardin, D. (1982). The assessment of binge eating severity among obese persons. Addictive Behaviors, 7, 47-55.

Graesser, A.C. (1981). Prose comprehension beyond the word. New York: Springer-Verlag.

Guyll, M., & Contrada, R.J. (1998). Trait hostility and ambulatory cardiovascular activity: Responses to social interaction. Health Psychology, 17, 30-39.

Henderson, M., & Freeman, C.P.L. (1987). A self-rating scale for bulimia: The BITE. British Journal of Psychiatry, 150, 18-24.

Herman, P.C., & Polivy, J. (1980). Restrained eating. In A.J. Stunkard (ed.), Obesity. Philadelphia: Saunders.

Israel, A.C., Silverman, W.K., & Solotar, L.C. (1998). The relationship between adherence and weight loss in a behavioral treatment program for overweight children. Behavior Therapy, 19, 25-33.

Jerome, A., Behar, A., & Dobbs, E. (1996). Using hand-held computers to implement self-management programs. Presented at the 30th Annual Convention of the Association for the Advancement of Behavior Therapy 1996. New York, NY.

Jerome, A., Frederiksen, L.W., Fredericksen, C.B. (1991). Computerized self-management of obesity. Presented at the 25th Annual Convention of the Association for the Advancement of Behavior Therapy, 1991. New York, NY.

Jobe, J.B. & Mingay, D.J. (1989). Cognitive research improves questionnaires. American Journal of Public Health, 79, 1053-1055.

Jobe, J.B., & Mingay, D.J. (1991). Cognition and survey measurement: History and overview. Applied Cognitive Psychology, 5, 175-192.

Jobe, J.B., Tourangeau, R., & Smith, A.F. (1994). Contributions of survey research to the understanding of memory. Applied Cognitive Psychology.

Johnson, R.K., Goran, M.I., & Poehlman, E.T. (1994). Correlates of over- and underreporting of energy intake in healthy older men and women. American Journal of Clinical Nutrition, 59, 1286-1290.

Johnson, C.L., & Larson, R. (1982). Bulimia: An analysis of moods and behavior. Psychosomatic Medicine, 44, 341-351.

Kahneman, D. (1973). Attention and effort. Englewood Cliffs, NJ: Prentice-Hall.

Kamarck, T.W., Shiffman, S.M., Smithline, L., Goodie, J.L., Paty, J.A., Gnys, M., & Jong, J.Y.K. (1998). The effects of task strain, social conflict, and emotional activation on ambulatory cardiovascular activity: Daily life consequences of recurring stress in a multiethnic adult sample. Health Psychology, 17, 17-29.

Lichtman, S.W., Pisarka, K., Berman, E.R., Pestone, M., Dowling, H., Offenbacher, E., Weisel, H., Heshka, S., Matthews, D.E., & Heymsfield, S.B. (1992). Discrepancy between self-reported and actual caloric intake and exercise in obese subjects. The New England Journal of Medicine, 327, 1893-1898.

Litt, M.D., Cooney, N.L., & Morse, P. (1998). Ecological momentary assessment (EMA) with treated alcoholics: Methodological problems and potential solutions. Health Psychology, 17, 48-52.

Loeb, K., Walsh, T.B., & Pike, K.M. (1992, April 24). The assessment of bulimia nervosa. Paper presented at the Fifth International Conference on Eating Disorders, New York.

Lowe, M.R., Kopyt, D., & Buchwald, J. (1996). Food intake underestimation: Its nature and potential impact on obesity treatment. The Behavior Therapist, 19(2), 17-20.

Lowe, M.R., Gleaves, D.H., DiSimone-Weiss, R.T., Furgueson, C., Gayda, C.A., Kolsky, P.A., Neal-Walden, T., Nelson, L.A., & McKinney, S. (1996). Restraint, dieting, and the continuum model of bulimia nervosa. Journal of Abnormal Psychology, 105, 508-517.

Marcus, M.D., Smith, D., Santelli, R., & Kaye, W. (1992). Characterization of eating disordered behavior in obese binge eaters. International Journal of Eating Disorders, 12, 249-256.

Means, B., Swan, G.E., Jobe, J.B., & Esposito, J.L. (1991). An alternative approach to obtaining personal history data. In P.P. Giemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, & S. Sudman (Eds.). Measurement errors in surveys. New York: Wiley and Sons.

Menon, G. (1993). The effects of accessibility of information in memory on judgments of behavioral frequencies. Journal of Consumer Research, 20, 431-440.

Metropolitan Life Insurance Company. (1983). Metropolitan height and weight standards. Statistical Bulletin of New York Metropolitan Life Insurance Company, 64, 2-9.

Mitchell, J.E., & Laine, D.C. (1985). Monitored binge eating behavior in patients with bulimia nervosa. International Journal of Eating Disorders, 4, 177-183.

Mitrany, E. (1992). Atypical eating disorders. Journal of Adolescent Health, 13, 400-402.

Okenfels, M.C., Porter, L., Smyth, J., Kirschbaum, C., Hellhammer, D.H., & Stone, A.A. (1995). The effect of chronic stress associated with unemployment on salivary cortisol: Overall cortisol levels, diurnal rhythm, and acute stress reactivity. Psychosomatic Medicine, 57, 460-467.

Orta, J., & Reinarts, C.L. (1994). Comparison of handheld computers for nutritional assessment and support. Journal of The American Dietetic Association, 94, 1409-14.

Palmer, R., Christie, M., Cordle, C., Davies, D., & Kendrick, J. (1987). The Clinical Eating Disorder Rating Instrument (CEDRI): A preliminary description. International Journal of Eating Disorders, 6, 9-16.

Peveler, R.C., Fairburn, C.G., Boller, I., & Dunger, D. (1992). Eating disorders in adolescents with insulin-dependent diabetes mellitus: A controlled study. Diabetes Care, 15, 1356-1360.

- Rabinowitz, J.C., Mandler, G., & Patterson, K.E. (1977). Determinants of recognition and recall: Accessibility and generation. Journal of Experimental Psychology, General, 106, 302-329.
- Reder, L.M. (1987a). Beyond associations: Strategic components in memory retrieval. In D.S. Gorfein and R.R. Hoffman (eds.). Memory and learning: The Ebbinghaus centennial conference (pp. 203-220). Hillsdale, NJ: Erlbaum.
- Reder, L.M. (1987b). Strategy selection in question answering. Cognitive Psychology, 19, 90-138.
- Roediger, H.L., & Payne, D.G. (1982). Hypermnnesia: The role of repeated testing. Journal of Experimental Psychology: Learning, memory, and Cognition, 8, 66-72.
- Rosen, J.C., Vara, L., Wendt, S., & Leitenberg, H. (1990). Validity studies of the Eating Disorder Examination. International Journal of Eating Disorders, 9, 519-528.
- Rosen, J.C., & Srebnik, D. (1990). Assessment of eating disorders. In P. McReynolds, J.C. Rosen, & G. Chelune (Eds.), Advances in psychological assessment (Vol. 7, pp. 229-259). New York: Plenum Press.
- Ross, L. (1984). Thoughts and research on estimates about past and future behavior. In T.M. Jabine, et al. (Eds.). Cognitive aspects of survey methodology: Building a bridge between disciplines. Washington, D.C.: National Academy Press.
- Rossiter, E.M., Agras, W.S., Telch, C.F., & Bruce, B. (1992). The eating patterns of non-purging bulimic subjects. International Journal of Eating Disorders, 11, 111-120.
- Sbrocco, T., Stone, J.M., Nedegaard, R.C., Lewis, E.L., Patel, K., & Gallant, L. (1997). Use of palmtop computers to assess dietary intake: A comparison of obese and normal weight women. Manuscript under review.
- Schlundt, D.G. (1993). Self-Monitoring Analysis System (Version 4.0) [Computer software]. Nashville, TN: Vanderbilt University.

- Schlundt, D.G., & Johnson, W.G., (1990). Eating disorders: Assessment and treatment. Boston: Allyn & Bacon.
- Schoeller, D.A. (1990). How accurate is self-reported dietary energy intake? Nutrition Reviews, 48, 373-379.
- Schoeller, D.A., Bandini, L.G., & Dietz, W.H. (1990). Inaccuracies in self-reported intake identified by comparison with the doubly labelled water method. Canadian Journal of Physiology and Pharmacology, 68, 941-949.
- Schwartz, J.E., & Stone, A.A. (1998). Strategies for analyzing ecological momentary assessment data. Health Psychology, 17, 6-16.
- Schwarz, N. (1990). Assessing frequency reports of mundane behavior: Contribution of cognitive psychology to questionnaire construction. In C. Hendrick & M.S. Clark (Eds.), Research methods in personality and social psychology. Newbury Park, CA: Sage.
- Shiffman, S., Fischer, L.A., Paty, J.A., Gnys, M., Kassel, J.D., Hickcox, M., & Perz, W. (1994). Drinking and smoking: A field study of their association. Annals of Behavioral Medicine, 16, 203-209.
- Shiffman, S., Paty, J.A., Gnys, M., Kassel, J.D., & Hickcox, M. (1996). First lapses to smoking: Within-subjects analysis of real-time reports. Journal of Consulting and Clinical Psychology, 64, 366-379.
- Shiffman, S., & Stone, A.A. (1998). Introduction to the special section: Ecological momentary assessment in health psychology. Health Psychology, 17, 3-5.
- Smith, A.B., Jobe, J.B., & Mingay, D.J. (1991). Question-induced cognitive biases in reports of dietary intake by college men and women. Health Psychology, 10, 244-251.
- Smith, M.C., & Thelan, M.H. (1984). Development and validation of a test for bulimia nervosa. Journal of Consulting and Clinical Psychology, 52, 863-872.

Smyth, J., Ockenfels, M.C., Gorin, A.A., Catley, D., Porter, L.S., Kirschbaum, C., Hellhammer, D.H., & Stone, A.A. (1997). Individual differences in the diurnal cycle of cortisol. Psychoneuroendocrinology, 22, 89-105.

Smyth, J., Soefer, M., Hurewitz, A., & Stone, A.A. (1997). The use of intensive momentary assessments to monitor asthmatics' response to a tape recorded relaxation intervention. Unpublished manuscript, State University of New York at Stony Brook.

Spitzer, R.L., Devlin, M., Walsh, B.T., Hasin, D., Wing, R., Marcus, M., Stunkard, A., Wadden, T., Yanovski, S., Agras, S., Mitchell, J., & Nonas, C. (1991). Binge Eating Disorder: To be or not to be in DSM-IV? International Journal of Eating Disorders, 10, 627-629.

Spitzer, R.L., Devlin, M., Walsh, B.T., Hasin, D., Wing, R., Marcus, M., Stunkard, A., Wadden, T., Yanovski, S., Agras, W.S., Mitchell, J., & Nonas, C. (1992). Binge eating disorder: A multisite field trial of the diagnostic criteria. International Journal of Eating Disorders, 11, 191-203.

Stone, A.A., Broderick, J.B., Porter, L., & Kaell, A.T. (1997). The experience of rheumatoid arthritis pain and fatigue: Examining momentary reports and correlates over one week. Arthritis Care and Research, 10, 185-193.

Stone, J.M., Nedegaard, R.C., & Sbrocco, T. (1997). Disordered eating among obese women. Paper presented at the Association for the Advancement of Behavior Therapy, Miami Beach, FL.

Stone, A.A., Schwartz, J.E., Neal, J.M., Shiffman, S., Marco, C.A., Hickcox, M., Paty, J., Porter, L.S., & Cruise, L.J. (in press). How accurate are current coping assessments? A comparison of momentary versus end-of-day reports of coping efforts. Journal of Personality and Social Psychology.

Stone, A.A., & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. Annals of Behavioral Medicine, 16, 199-202.

Stone, A.A., Shiffman, S., & deVries, M. (in press). Rethinking self-report assessment methodologies: An argument for collecting ecologically valid, momentary measurements and selected results of EMA studies. In d. Kahneman, E. Diener, & N. Schwartz (Eds.), Understanding quality of life: Scientific perspectives on enjoyment and suffering. New York: Russell Sage.

Striegel-Moore, R.H., Nicholas, T.J., & Tamborlane, W.V. (1992). Prevalence of eating disorder symptoms in preadolescent and adolescent girls with IDDM. Diabetes Care, 15, 1361-1368.

Strube, G. (1987). Answering survey questions: The role of memory. In H.J. Hippler, et al. (eds.), Social information processing and survey methodology. New York: Springer.

Stunkard, A.J., & Messick, S. (1985). The Three-Factor Eating Questionnaire to measure dietary restraint, disinhibition, and hunger. Journal of Psychosomatic Research, 29, 71-83.

Taylor, A.V., Peveler, R.C., Hibbert, G.A., & Fairburn, C.G. (1994). Eating disorders among women receiving treatment for an alcohol problem. International Journal of Eating Disorders.

Tulving, E., & Thomson, D.M. (1973). Encoding specificity and retrieval processes in episodic memory. Psychological Review, 80, 352-373.

Van Strien, T., Frijters, J.E.R., Bergers, G.P.A., & Defares, P.B. (1986). The Dutch Eating Behavior Questionnaire (DEBQ) for assessment of restrained, emotional, and external eating behavior. International Journal of Eating Disorders, 5, 295-315.

Vuchinich, R.E., Tucker, J.A., & Harllee, L.M. (1988). Behavioral assessment. In D.M. Donovan & G.A. Marlatt (Eds.), Assessment of addictive behaviors (pp. 51-83). New York: Guilford Press.

Wadden, T.A., & Foster, G.D. (1992). Behavioral assessment and treatment of markedly obese patients. In T.A. Wadden & T.B. VanItallie (Eds.), Treatment of the seriously obese patient (pp. 290-330). New York: Guilford Press.

Williamson, D.A. (1990). Assessment of eating disorders. New York: Pergamon Press.

Williamson, D.A., Anderson, D.A., & Gleaves, D.H. (1996). Anorexia and bulimia nervosa: Structured interview methodologies and psychological assessment. In K. Thompson (Ed.), Body image, eating disorders, and obesity: A practical guide for assessment and treatment (pp. 205-223). Washington, D.C.: American Psychological Association.

Willis, G.B., Royston, P., & Bercini, D. (1991). The use of verbal report methods in the development and testing of survey questionnaires. Applied Cognitive Psychology, 5, 251-267.

Wilson, G.T. (1987). Assessing treatment outcome in bulimia nervosa: A methodological note. International Journal of Eating Disorders, 6, 339-348.

Wilson, G.T. (1993). Behavioral treatment of obesity: Thirty years and counting. Advances in Behaviour Research and Therapy, 16, 31-75.

Wilson, G.T., & Smith, D. (1989). Assessment of bulimia nervosa and evaluation of the Eating Disorder Examination. International Journal of Eating Disorders, 8, 173-180.

Wilson, G.T., Eldredge, K.L., Smith, D., & Niles, B. (1991). Cognitive-behavioral treatment with and without response prevention for bulimia. Behaviour Research and Therapy, 29, 575-583.

TABLES

Table 1

DSM-IV Diagnostic Criteria for Bulimia Nervosa

- A. Recurrent episodes of binge eating. An episode of binge eating is characterized by both of the following:
- (1) eating, in a discrete period of time (e.g., within any 2-hour period), an amount of food that is definitely larger than most people would eat during a similar period of time and under similar circumstances.
 - (2) a sense of lack of control over eating during the episode (e.g., a feeling that one cannot stop eating or control what or how much one is eating).
- B. Recurrent inappropriate compensatory behavior in order to prevent weight gain, such as self-induced vomiting; misuse of laxatives, diuretics, enemas, or other medications; fasting; or excessive exercise.
- C. The binge eating and inappropriate compensatory behaviors both occur, on average, at least twice a week for 3 months.
- D. Self-evaluation is unduly influenced by body shape and weight.
- E. The disturbance does not occur exclusively during episodes of Anorexia Nervosa.

Specify type:

Purging Type: during the current episode of Bulimia Nervosa, the person has regularly engaged in self-induced vomiting or the misuse of laxatives, diuretics, or enemas.

Nonpurging Type: during the current episode of Bulimia Nervosa, the person has used other inappropriate compensatory behaviors, such as fasting or excessive exercise, but has not regularly engaged in self-induced vomiting or the misuse of laxatives, diuretics, or enemas.

Note. From Diagnostic and statistical manual of mental disorders (4th ed.) (pp. 549 – 550), by the American Psychiatric Association, 1994, Washington, DC: Author.

Table 2

DSM-IV Diagnostic Criteria for Anorexia Nervosa

- A. Refusal to maintain body weight at or above a minimally normal weight for age and height (e.g., weight loss leading to maintenance of body weight less than 85% of that expected; or failure to make expected weight gain during period of growth, leading to body weight less than 85% of that expected).
- B. Intense fear of gaining weight or becoming fat, even though underweight.
- C. Disturbance in the way in which one's body weight or shape is experienced, undue influence of body weight or shape on self-evaluation, or denial of the seriousness of the current low body weight.
- D. In post-menarchal females, amenorrhea, i.e., the absence of at least three consecutive menstrual cycles. (A woman is considered to have amenorrhea if her periods occur only following hormone, e.g., estrogen, administration.)

Specify type:

Restricting Type: during the current episode of Anorexia Nervosa, the person has not regularly engaged in binge-eating or purging behavior (i.e., self-induced vomiting or the misuse of laxatives, diuretics, or enemas).

Binge-Eating/Purging Type: during the current episode of Anorexia Nervosa, the person has regularly engaged in binge-eating or purging behavior (i.e., self-induced vomiting or the misuse of laxatives, diuretics, or enemas).

Note. From Diagnostic and statistical manual of mental disorders (4th ed.) (pp. 544 – 545), by the American Psychiatric Association, 1994, Washington, DC: Author.

Table 3

DSM-IV Eating Disorder Not Otherwise Specified

The Eating Disorder Not Otherwise Specified category is for disorders of eating that do not meet the criteria for any specific Eating Disorder. Examples include:

1. For females, all of the criteria for Anorexia Nervosa are met except that the individual has regular menses.
 2. All of the criteria for Anorexia Nervosa are met except that, despite significant weight loss, the individual's current weight is in the normal range.
 3. All of the criteria for Bulimia Nervosa are met except that the binge eating and inappropriate compensatory mechanisms occur at a frequency of less than twice a week or for a duration of less than 3 months.
 4. The regular use of inappropriate compensatory behavior by an individual or normal body weight after eating small amounts of food (e.g., self-induced vomiting after the consumption of two cookies).
 5. Repeatedly chewing and spitting out, but not swallowing, large amounts of food.
 6. Binge-eating disorder.
-

Note. From Diagnostic and statistical manual of mental disorders (4th ed.) (p. 550), by the American Psychiatric Association, 1994, Washington, DC: Author.

Table 4

DSM-IV Research Criteria for Binge Eating Disorder

A. Recurrent episodes of binge eating. An episode of binge eating is characterized by both of the following:

- (1) eating, in a discrete period of time (e.g., within any 2-hour period), an amount of food that is definitely larger than most people would eat in a similar period of time under similar circumstances.**
- (2) a sense of lack of control over eating during the episode (e.g., a feeling that one cannot stop eating or control what or how much one is eating).**

B. The binge-eating episodes are associated with three (or more) of the following:

- (1) eating much more rapidly than normal**
- (2) eating until feeling uncomfortably full**
- (3) eating large amounts of food when not feeling physically hungry**
- (4) eating alone because of being embarrassed by how much one is eating**
- (5) feeling disgusted with oneself, depressed, or very guilty after overeating**

C. Marked distress regarding binge eating is present.

D. The binge eating occurs, on average, at least 2 days a week for 6 months.

Note: The method of determining frequency differs from that used for Bulimia Nervosa; future research should address whether the preferred method of setting a frequency threshold is counting the number of days on which binges occur or counting the number of episodes of binge eating.

E. The binge eating is not associated with the regular use of inappropriate compensatory behaviors (e.g., purging, fasting, excessive exercise) and does not occur exclusively during the course of Anorexia Nervosa or Bulimia Nervosa.

Note. From Diagnostic and statistical manual of mental disorders (4th ed.) (p. 731), by the American Psychiatric Association, 1994, Washington, DC: Author.

Table 5

The Four Subscales of EDE 12.0D

Restraint	Shape Concern
Restraint over eating	Flat stomach
Avoidance of eating	Importance of shape
Food avoidance	Preoccupation with shape or weight
Dietary Rules	Dissatisfaction with shape
Empty stomach	Fear of weight gain
	Discomfort seeing body
	Avoidance of exposure
	Feeling of fatness
Eating Concern	Weight Concern
Preoccupation with food, eating, or calories	Importance of weight
Fear of losing control over eating	Reaction to prescribed weighing
Social eating	Preoccupation with shape or weight
Eating in secret	Dissatisfaction with weight
Guilt about eating	Desire to lose weight

Note. From “The Eating Disorder Examination (12th Edition)” by C.G. Fairburn, and Z. Cooper, 1993, in C.G. Fairburn & G.T. Wilson (Eds.), Binge eating: Nature, assessment, and treatment, p. 319.

Table 6

The EDE Rating Scheme

Severity ratings	Frequency ratings
0 – Absence of the feature	0 – Absence of the feature
1 – Feature almost, but not quite, absent	1 – Feature present on 1-5 days
2 –	2 – Feature present on 6-12 days
3 – Severity midway between 0 and 6	3 – Feature present on 13-15 days
4 –	4 – Feature present on 16-22 days
5 – Feature present to a degree not quite severe enough to justify a rating of 6	5 – Feature present almost every day (23-27 days)
6 – Feature present to an extreme degree	6 – Feature present every day

Note. From “The Eating Disorder Examination (12th Edition)” by C.G. Fairburn, and Z. Cooper, 1993, in C.G. Fairburn & G.T. Wilson (Eds.), Binge eating: Nature, assessment, and treatment, p. 330.

Table 7

Obese and Normal Weight Participant Characteristics

	Obese (n=17) Mean (SD)	Normal Weight (n=15) Mean (SD)
Age	39.3 (9.8) _a	39.3 (11.3) _a
Body Weight (kg)	87.9 (10.6) _a	58.6 (6.2) _b
Body Mass Index (BMI) (kg/m²)	32.2 (3.6) _a	22.2 (2.6) _b

Note. Means in the same row that do not share subscripts differ at $p < .05$.

Table 8

Comparisons of reported meal frequencies for the Eating Disorder Examination (EDE) and handheld computer eating diary

Meal	EDE (0 – 6) (SD)	Diary (0 – 6) (SD)	Diary (%) (SD)	Kendall's tau-b	Wilcoxon Matched Pairs z score
All Subjects (n=32)					
breakfast	4.58 (1.68)	5.09 (1.06)	86.88 (18.34)	0.27*	-1.78 ^{a*}
morning snack	2.14 (2.13)	1.09 (1.15)	17.09 (19.71)	0.31*	-2.37 ^{b**}
lunch	5.53 (0.88)	4.97 (0.93)	84.86 (17.10)	0.08	-1.74 ^{b*}
afternoon snack	2.47 (2.41)	2.78 (1.48)	45.77 (28.30)	0.31*	-0.59 ^a
dinner	5.42 (1.20)	4.97 (0.93)	84.19 (15.61)	0.50**	-2.28 ^{b**}
evening snack	2.08 (1.95)	2.03 (1.53)	34.61 (30.73)	0.38**	-1.46 ^{a*}
Obese (n=17)					
breakfast	4.60 (1.60)	5.41 (0.94)	90.59 (15.69)	0.27	-2.09 ^{***}
morning snack	2.30 (2.32)	1.0 (1.27)	16.29 (21.36)	0.31	-1.29 ^{b*}
lunch	5.80 (0.41)	4.88 (0.93)	83.49 (15.86)	0.02	-2.33 ^{b**}
afternoon snack	2.90 (2.53)	3.00 (1.41)	50.10 (25.95)	0.18	-0.04 ^a
dinner	5.70 (1.13)	5.29 (0.92)	89.04 (16.06)	0.35	-1.06 ^b
evening snack	2.75 (2.24)	2.65 (1.58)	45.70 (34.73)	0.40*	-1.37 ^{a*}
Normals (n=15)					
breakfast	4.56 (1.82)	4.73 (1.10)	82.67 (20.67)	0.48**	-0.09 ^a
morning snack	1.94 (1.91)	1.20 (1.01)	17.99 (18.36)	0.42*	-2.08 ^{b**}
lunch	5.19 (1.17)	5.07 (0.96)	86.42 (18.84)	0.13	-0.26 ^b
afternoon snack	1.94 (2.21)	2.53 (1.55)	40.86 (30.90)	0.43*	-0.91 ^a
dinner	5.06 (1.24)	4.60 (0.83)	78.71 (13.57)	0.50**	-2.14 ^{b**}
evening snack	1.25 (1.06)	1.33 (1.18)	22.05 (19.86)	0.31*	-0.32 ^a

Note. The following EDE rating scheme was used: 0 – meal or snack not eaten, 2 – meal or snack eaten on less than half the days, 4 – meal or snack eaten on more than half the days, 6 – meal or snack eaten every day.

^a Diary > EDE. ^b Diary < EDE.

* $p < .10$. ** $p < .05$.

Table 9

Comparisons of the Eating Disorder Examination (EDE) and handheld computer eating diary assessment of restraint over eating

	EDE (0 – 6) (SD)	Diary (0 – 6) (SD)	Diary (% days kcal < 2SD) (SD)	Kendall's tau-b	Wilcoxon Matched Pairs z score
All Subjects (n=32)	2.19 (2.61)	1.00 (0.86)	13.57 (12.63)	0.18	-0.64 ^b
Obese (n=17)	3.40 (2.50)	1.12 (0.86)	14.71 (12.74)	0.02	-1.79 ^{b*}
Normals (n=15)	0.69 (1.89)	0.86 (0.86)	12.09 (12.84)	0.11	-1.65 ^{a*}

Note. The following EDE rating scheme was used: 0 – no attempt at restraint, 2 – attempted to exercise restraint on less than half the days, 4 – attempted to exercise restraint on more than half the days, 6 – attempted to exercise restraint every day.

^a Diary > EDE. ^b Diary < EDE.

* p < .10.

Table 10

Comparisons of the Eating Disorder Examination (EDE) and handheld computer eating diary assessment of avoidance of eating

	EDE (0 – 6) (SD)	Diary (0 – 6) (SD)	Diary (% days w/ 8 hrs between meals) (SD)	Kendall's tau-b	Wilcoxon Matched Pairs z score
All Subjects (n=32)	0.03 (0.17)	1.28 (1.02)	21.02 (17.55)	0.27*	-4.00***
Obese (n=17)	0.05 (0.22)	1.12 (1.05)	19.31 (17.83)	0.42*	-2.40***
Normals (n=15)	0.00 (0.00)	1.47 (0.99)	22.95 (17.63)	c	-3.25***

Note. The following EDE rating scheme was used: 0 – no such days, 2 – avoidance on less than half the days, 4 – avoidance on more than half the days, 6 – avoidance every day.

^a Diary > EDE. ^b Diary < EDE. ^c Cannot calculate because no range of EDE scores.

* p < .10. **p < .05.

Table 11

Comparisons of the Eating Disorder Examination (EDE) and handheld computer eating diary
assessment of days of overeating

	EDE (# days) (SD)	Diary (# days kcal > 2SD) (SD)	Mean kcal/day (SD)	Kendall's tau-b	Wilcoxon Matched Pairs z score
All Subjects (n=32)	3.36 (5.09)	3.68 (3.34)	2291 (615)	0.11	-0.85 ^a
Obese (n=17)	6.05 (5.53)	3.65 (3.23)	2462 (550)	0.16	-1.53 ^{b*}
Normals (n=15)	0.00 (0.00)	3.73 (3.61)	2029 (634)	c	-2.82 ^{***}

^a Diary > EDE. ^b Diary < EDE. ^c Cannot calculate because no range of EDE scores.

* p < .10. **p < .05.

Table 12

Comparisons of the Eating Disorder Examination (EDE) and handheld computer eating diary
assessment of episodes of overeating

	EDE (# episodes) (SD)	Diary (# meals > 1000 kcal) (SD)	Diary (# meals "overeaten") (SD)	Kendall's tau-b	Wilcoxon Matched Pairs z score
All Subjects (n=32)	4.00 (6.21)	17.74 (11.95)	38.71 (45.26)	0.16, 0.32*	-4.35 ^{b**} , -3.94 ^{a**}
Obese (n=17)	7.20 (6.83)	18.12 (11.16)	52.37 (54.19)	0.23, 0.08	-2.85 ^{b**} , -2.85 ^{a**}
Normals (n=15)	0.00 (0.00)	17.33 (13.13)	20.50 (19.78)	c, c	-3.30 ^{b**} , -2.80 ^{a**}

^a Diary > EDE. ^b Diary < EDE. ^c Cannot calculate because no range of EDE scores.

* p < .10. **p < .05.

Table 13

Assessment of episodes of overeating from the handheld computer eating diary

	Mean kcal/meal (SD)	kcal/meal when "overeaten" (SD)	kcal/meal when "not overeaten" (SD)	% meals "overeaten" (SD)
All Subjects (n=32)	620 (270)	703 (459)	461 (229)	37 (31)
Obese (n=17)	553 (128)	766 (500)	388 (218)	45 (34)
Normals (n=15)	690 (358)	624 (408)	552 (218)	26 (24)

Table 14

Comparisons of the Eating Disorder Examination (EDE) and handheld computer eating diaryassessment of reaction to prescribed weighing

	EDE (0 - 6) (SD)	Diary (0 - 6) (SD)	Diary (early daily kcal – late daily kcal) (SD)	Kendall's tau-b	Wilcoxon Matched Pairs z score
All Subjects	0.46 (1.17)	1.56 (1.56)	157.09 (411.25)	-0.06	-2.93 ^{***}
Obese	0.79 (1.51)	1.56 (1.72)	122.61 (448.95)	-0.13	-1.48 ^{**}
Normals	0.06 (0.25)	1.57 (1.40)	201.43 (368.75)	-0.07	-2.84 ^{***}

Note. The following EDE rating scheme was used: 0 – no reaction, 2 – slight reaction, 4 – moderate reaction, 6 – marked reaction. The following diary rating scheme was used: 0 – Kcal of last four days not less than first ten days, 1 – difference of 167 kcal, 2 – difference of 333 kcal, 3 – difference of 500 kcal, 4 – difference of 667 kcal, 5 – difference of 833 kcal, 6 – difference of 1000 kcal.

^a Diary > EDE. ^b Diary < EDE.

* p < .10. **p < .05.

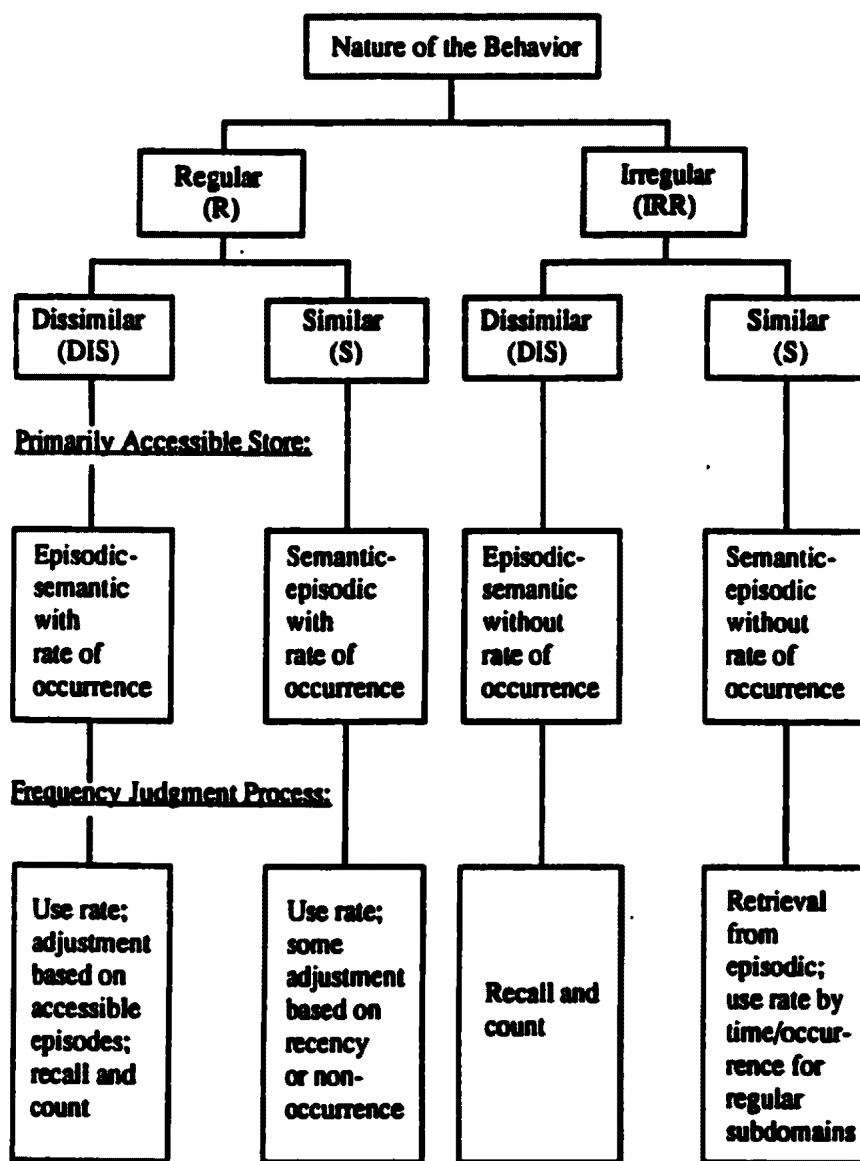


Figure 1. Proposed model for the storage and retrieval of information about frequent behavior.

“Semantic-episodic” implies a more accessible semantic store with few accessible episodes, whereas “episodic-semantic” implies more accessible episodes with/without accessible semantic stores with subdomains. From “The effects of accessibility of information in memory on judgments of behavioral frequencies,” by G. Menon, 1993, Journal of Consumer Research, 20, p. 435.

	Amount Eaten	
	"large" (EDE definition)	Not "large," but viewed by subject as excessive
"Loss of Control"	Objective bulimic episodes	Subjective bulimic episodes
No "loss of control"	Objective overeating	Subjective overeating

Figure 2. The EDE Scheme for classifying episodes of overeating. From "The Eating Disorder Examination (12th Edition)" by C.G. Fairburn, and Z. Cooper, 1993, in C.G. Fairburn & G.T. Wilson (Eds.), Binge eating: Nature, assessment, and treatment, p. 319.

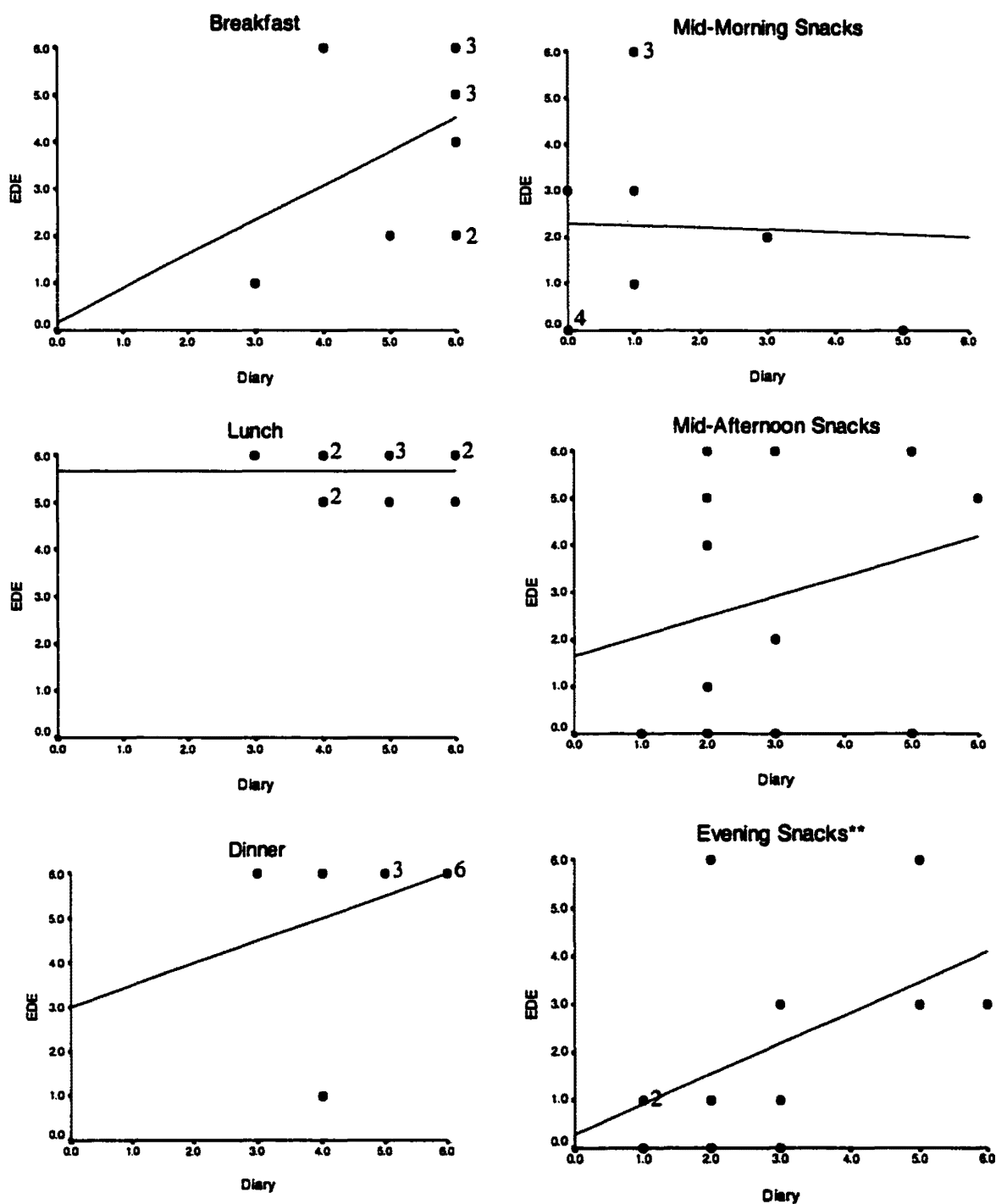


Figure 3. Comparisons of obese' responses for each meal on the EDE vs. the diary. The line represents the linear regression fit. Overlapping data points are represented by a number indicating the number of multiple responses at that point.

****Statistical trend ($p < .10$) for correlation between EDE and diary.**

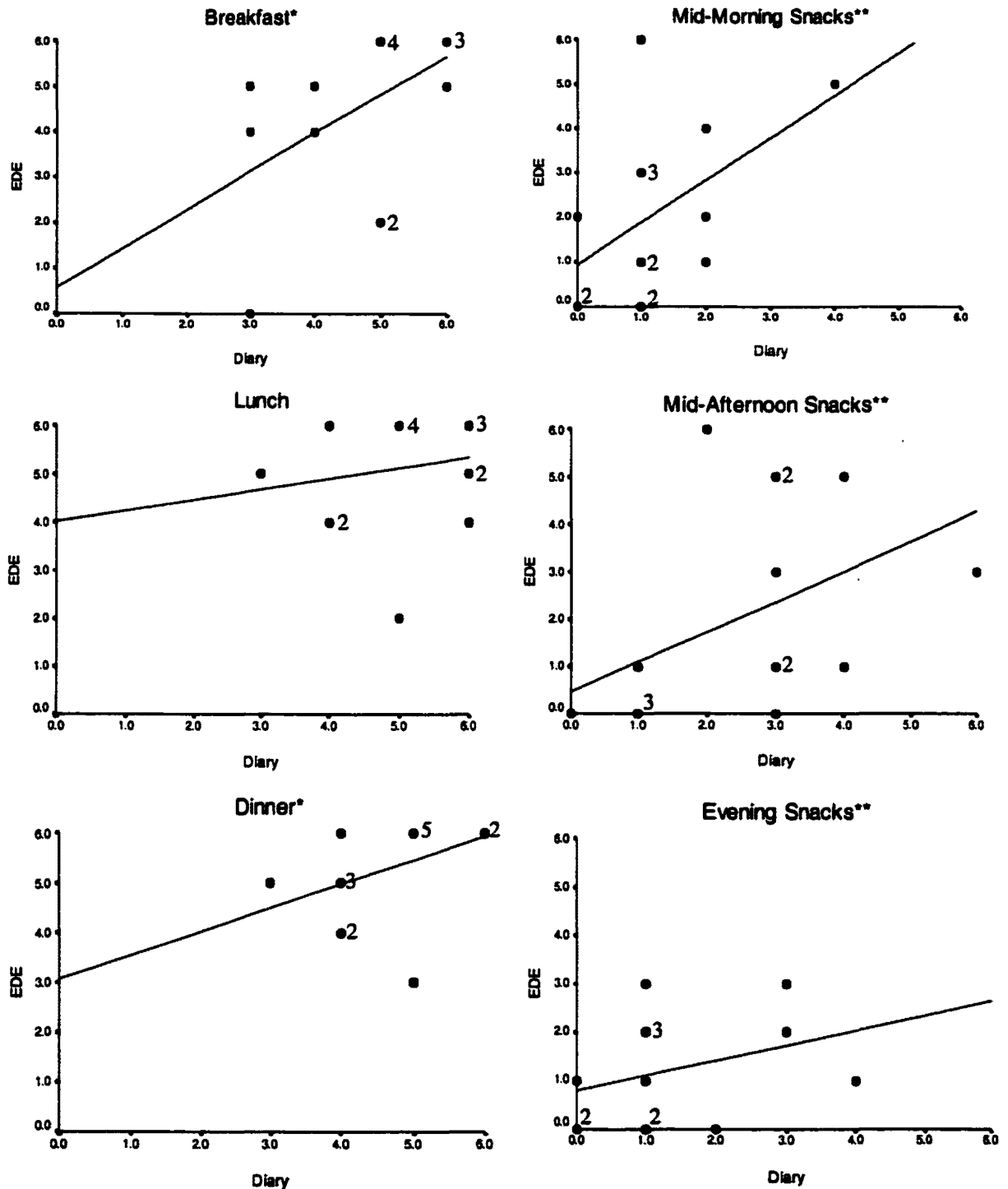


Figure 4. Comparisons of normals' responses for each meal on the EDE vs. the diary. The line represents the linear regression fit. Overlapping data points are represented by a number indicating the number of multiple responses at that point.

*Significant correlation ($p < .05$) between EDE and diary.

**Statistical trend ($p < .10$) for correlation between EDE and diary.

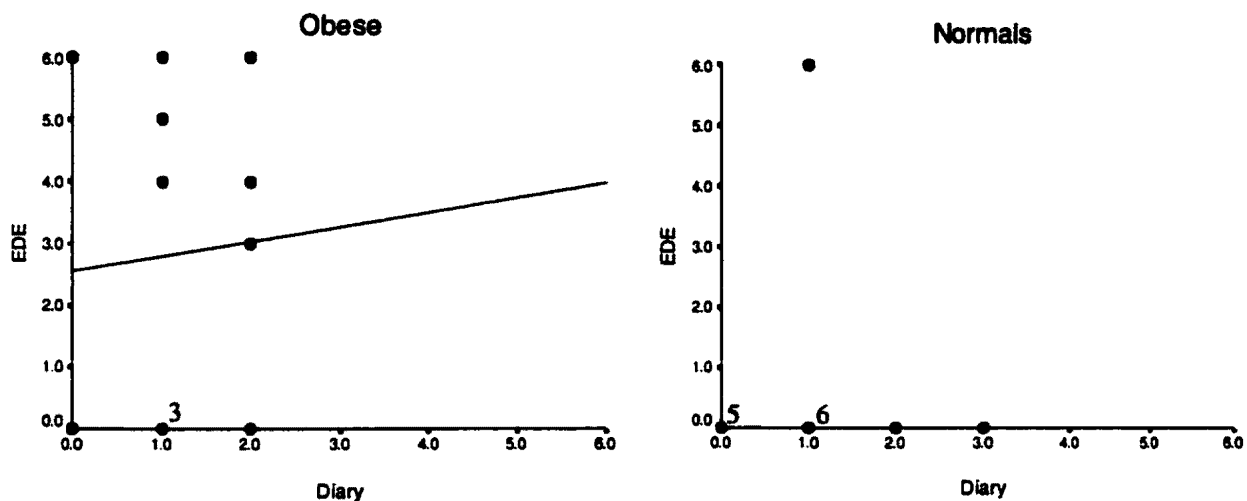


Figure 5. Comparisons of responses for restraint over eating on the EDE vs. the diary. The line represents the linear regression fit (when it can be calculated). Overlapping data points are represented by a number indicating the number of multiple responses at that point.

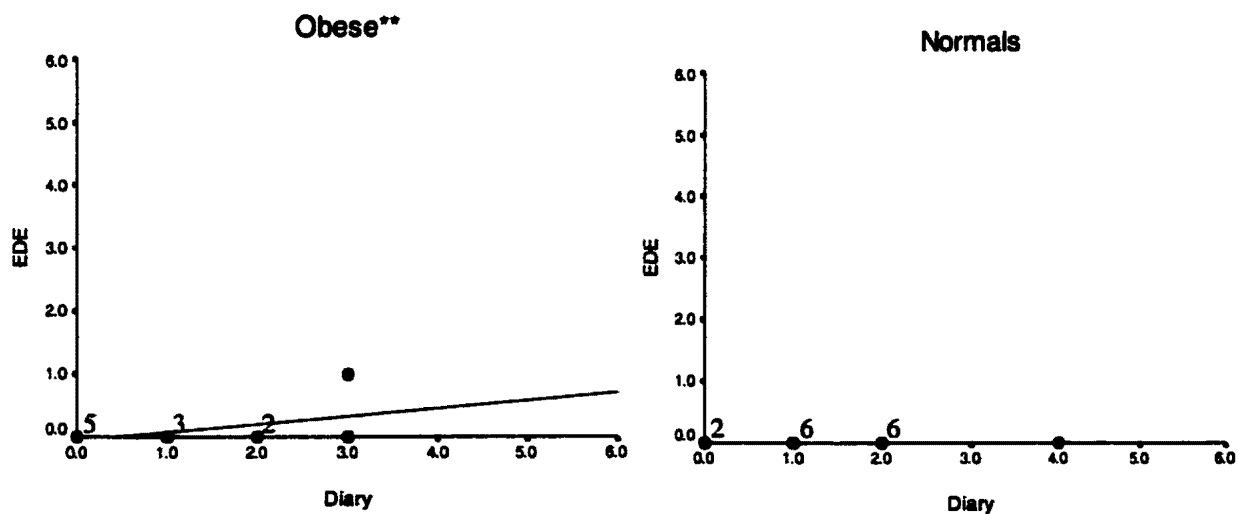


Figure 6. Comparisons of responses for avoidance of eating on the EDE vs. the diary. The line represents the linear regression fit (when it can be calculated). Overlapping data points are represented by a number indicating the number of multiple responses at that point.

****Statistical trend ($p < .10$) for correlation between EDE and diary.**

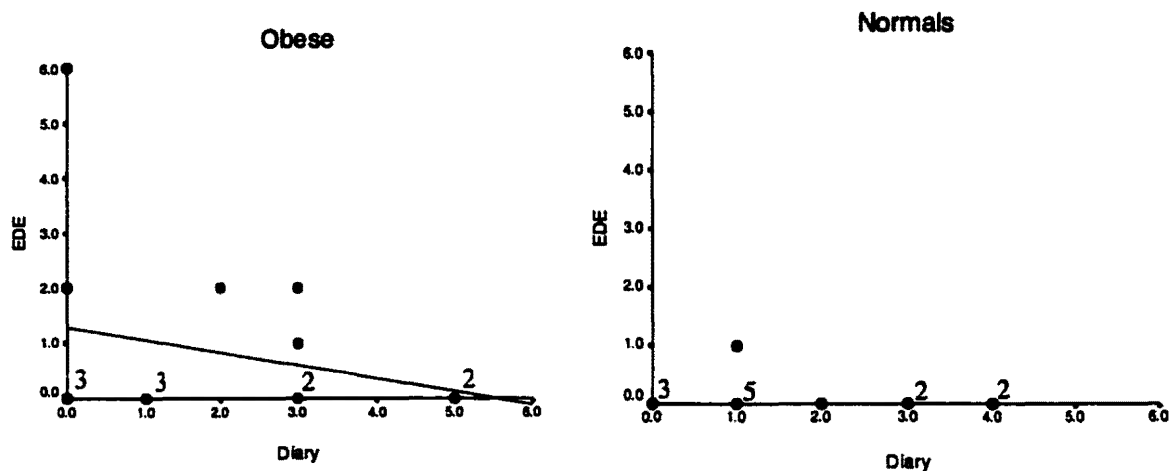


Figure 7. Comparisons of reaction to prescribed weighing on the EDE vs. the diary. The line represents the linear regression fit (when it can be calculated). Overlapping data points are represented by a number indicating the number of multiple responses at that point.